Software Analyst
Cyber Research

Data Security Platforms: The New Frontier In Cybersecurity & AI

The Data-Centric Revolution: How Data Security Is The Epic Center For The AI Era

Software Analyst
Cyber Research

Data Security Platforms: The New Frontier in Cybersecurity & AI

The Data-Centric Revolution: How Data Security Is The Epic Center for the AI Era

.Exploring the processes, challenges, solutions, and path toward a future of AI-Augmented Security Operations Centers (SOC)

By Francis Odum

Readers,

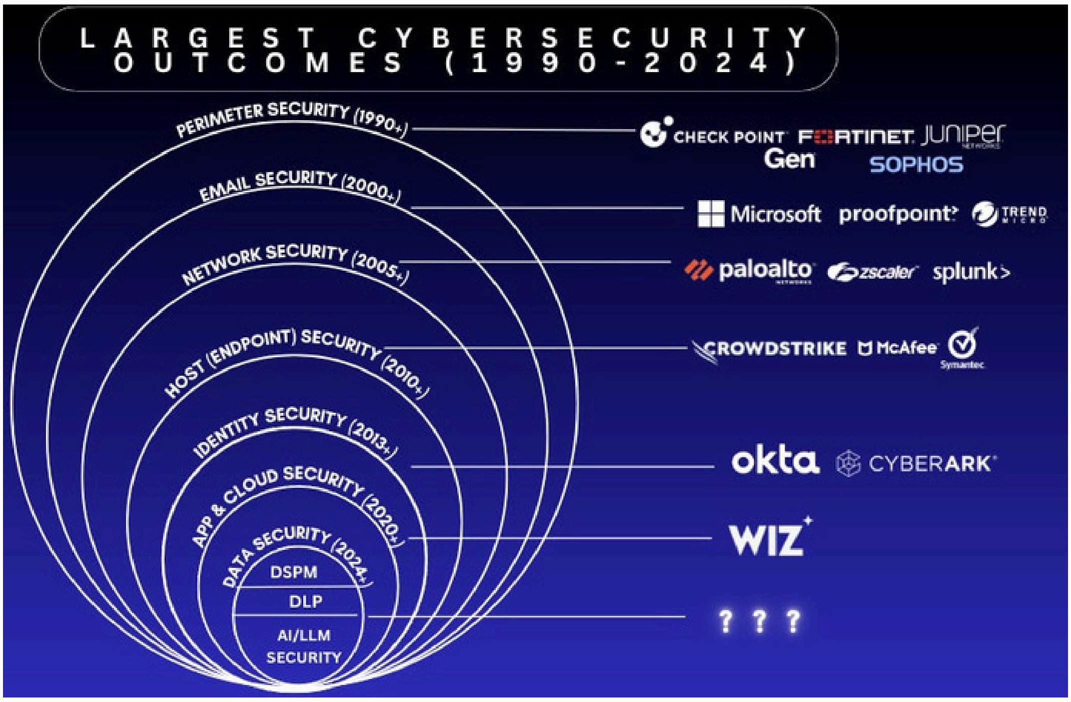
The data security landscape is undergoing a significant transformation and has reached an inflection point. This shift is driven by the proliferation of data, the rise of generative AI, and the increasing stringency of data compliance regulations. Traditional network-based security measures are no longer sufficient, as data itself has become the new perimeter. This necessitates a shift towards data-centric security strategies.

Organizations face numerous challenges, including data breaches, compliance penalties, and the dynamic nature of data itself. Other pressing issues involve the complexities of GenAI, cyber resilience, and data sprawl. To address these challenges, a comprehensive data security program must encompass data discovery, classification, governance, data protection, monitoring, response, and recovery.

I present a case for why the next big opportunity in cybersecurity will be focused in and around data which is the crown jewel for every organization. I collaborated with Cyera to generate the second half of the report, using their platform as a case study for the revolution across data security and how their platform illustrates this direction. Additionally, the research was built on extensive data security market research with different security operators. The Software Analyst spoke to a number of Chief Data Officers, leaders and security leaders to generate this report.

Data Security Is The Pinnacle of the Next Big Thing In Cybersecurity

If the history of cybersecurity is any indication of the future, then it likely leads to data security. We want to evaluate success - below, or the “next big thing“ by reviewing $1B+ in revenue companies (or close to $1B over the next 12-months) that have evolved within market categories using the layered security approach, which is a popular framework utilized by many security leaders. This timeline follows the history of cybersecurity to modern data security:



- **Perimeter Security (1990s):** The risks in the early days of the internet led to security vendors like Checkpoint and Juniper Networks, which were built to secure enterprise perimeters.
- **Network Security (2000s):** This era gave rise to companies like Palo Alto Networks and Zscaler, that were built around firewalls.
- **Email Security (2005):** We subsequently saw the rise of Proofpoint and Microsoft Security to protect this attack vector.
- **Host (Endpoint) Security (2010s):** We saw the rise of the importance of securing the operating system and endpoints with the initial success of McAfee, Symantec, TrendMicro, Sophos and eventually Crowdstrike.
- **Identity Security (2015’s)**: The rise of SaaS applications gave rise to the need for identity security controls to determine the right users and their access to enterprise resources. We saw Okta and subsequently, CyberArk emerge as a successful platform.
- **Cloud & Application Security (2020s):** Although cloud security always existed, the pandemic gave it a resurgence with the success of Wiz.
- **Data Security (2024?):** It leads us to ask why we’ve never seen the first billion-dollar data security platform.

The Case for a Holistic Data Security Platform

In the evolving landscape of cybersecurity, data security presents the greatest challenge. Despite being the most crucial asset, data often has the least robust controls. The proliferation of third-party solutions, cloud environments, SaaS, and AI has dispersed data across dynamic locations, including hybrid and multi-cloud systems. This, coupled with the widespread access by employees, contractors, and partners, creates significant complexities in visibility and control. As organizations increasingly adopt cloud-based and AI-driven architectures, where data is central, a fundamental shift in security approaches is imperative. Traditional data security solutions are no longer sufficient. The dynamic nature of modern data environments necessitates real-time, adaptable security measures. The time has come for a comprehensive, holistic data security platform that addresses these evolving challenges.

This Research Report

This report examines the data security world. It offers insights into the essential elements of a comprehensive data security program, including data discovery, classification, protection, and destruction. The report also emphasizes the role of AI/ML and the transition of legacy technologies in transforming data security initiatives. Additionally, it presents an overview of the competitive landscape and demonstrates how Cyera, a prominent data security platform, tackles these challenges through its holistic solutions. In conclusion, this report provides valuable knowledge for organizations striving to navigate the complexities of data security and protect their most valuable asset — data.

Key Actionable Takeaways

1. Security leaders are prioritizing data security projects into the new year. According to a YL Venture survey based on Forrester's "The State of Data Security" (July 2024), 83% of enterprises currently use endpoint DLP, but only 13% have fully deployed their data security capabilities in the cloud. This gap, combined with growing privacy and compliance demands, positions next-generation DLP solutions as critical components of future security strategies.
2. Based on a survey of 218+ CISOs, companies going into 2025 are budgeting to prioritize Data Security (including DLP), into their plans. The data suggests DLP is back as nearly half of data security projects involve data loss prevention, or DLP. This perennial market seems to be blooming under the possibilities that AI brings to the classification side of the problem; and while data security posture management as a separate category seems to be dwindling, secrets management, data vaulting and tokenization are all showing up as projects in conversations.
3. Gartner has identified data security around GenAI as the #1 cybersecurity trend for 2025, while #4 highlights organizations' growing desire to transition to cyber resilience. According to research by Cybersecurity insiders and Cyera, 75% of enterprises plan to adopt a DSPM solution over the next 12-months
4. Organizations are currently facing challenges around managing and using data across their enterprise. There is a lack of cross-functional collaboration across the enterprise. There is data across all aspects of the enterprise making it very difficult to identify and put together a good cross-functional strategy across the enterprise.
5. Every company will need a data security strategy to meet the demand and needs of a data centric world. Relative to other categories in the history of cybersecurity, data security has the least amount of developed security solutions. This presents data security with the opportunity to deliver the next billion-dollar company in cybersecurity.
6. Based on my research, the next data security platform will be built on strong data discovery and classification (DSPM) and holistic data protection mechanism built around Data Loss Prevention (DLP) as the core foundation to power other data related solutions in the enterprise.
7. Muji at the hypergrowth has a full analysis titled, The resiliency landscape of the intersection between cyber resiliency and data protection with cybersecurity CNAPP vendors like CrowdStrike, Palo Alto, and Zscaler.
8. This report aims to provide a detailed breakdown of the data security ecosystem and I've collaborated on this report with Cyera to show how companies like Cyera are leading the next generation of data security platforms.

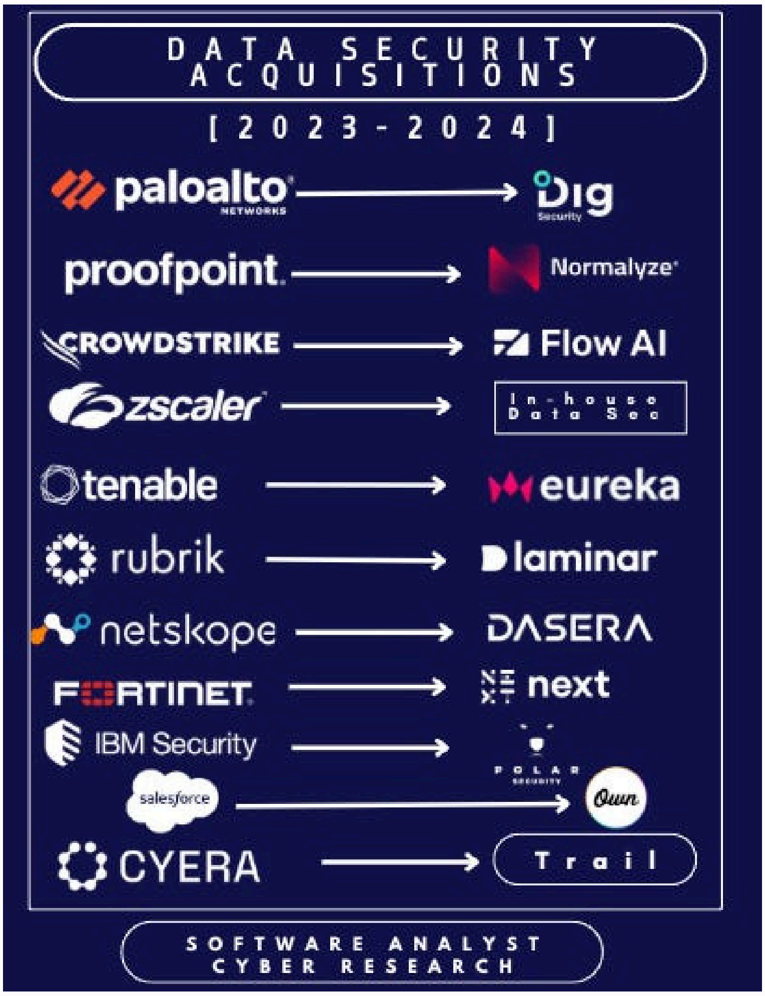
Proof: Surge in Data Security Acquisitions and Investments Suggests A Fight for Data

The largest cybersecurity companies have been in an arms race, trying to catch up by acquiring companies as quickly as they can to secure the crown jewel of data. The rising activity in data security investment and acquisitions underscores the growing demand for effective data protection solutions. The DSPM market has experienced substantial growth, highlighted by over $1 billion in funding and M&A transactions. For instance, Tenable's recent acquisition of Eureka Security demonstrates the increasing importance of data security as a key area of focus beyond its traditional vulnerability management solutions.

Significant acquisitions from major cybersecurity companies support this trend. For example,

- Palo Alto Networks acquired Dig Security
- CrowdStrike acquired Flow
- Rubrik acquired Laminar
- Proofpoint acquired Normalyze (2024)
- Netskope acquired Dasera (2024)
- Tenable acquired Eureka (2024)
- Cyera acquired Trail Security (2024) and secured $300 million at a $3B valuation.

I wrote more about the state of these acquisitions in a detailed LinkedIn post here:



According to Altitude Cyber, the volume of deals is also on the rise, with 114 deals totaling nearly $2.6 billion in 2024 alone. Since 2020, there have been 576 deals totaling $9.9 billion in funding, illustrating the critical role data security plays in the ongoing development of cybersecurity solutions. Additionally, venture capital investment in data security companies has been substantial, with over $500 million raised across the sector. These high-profile transactions reflect the broader industry trend where established cybersecurity firms are investing in data security. The data security market continues to grow and we expect it to grow over the next few years.

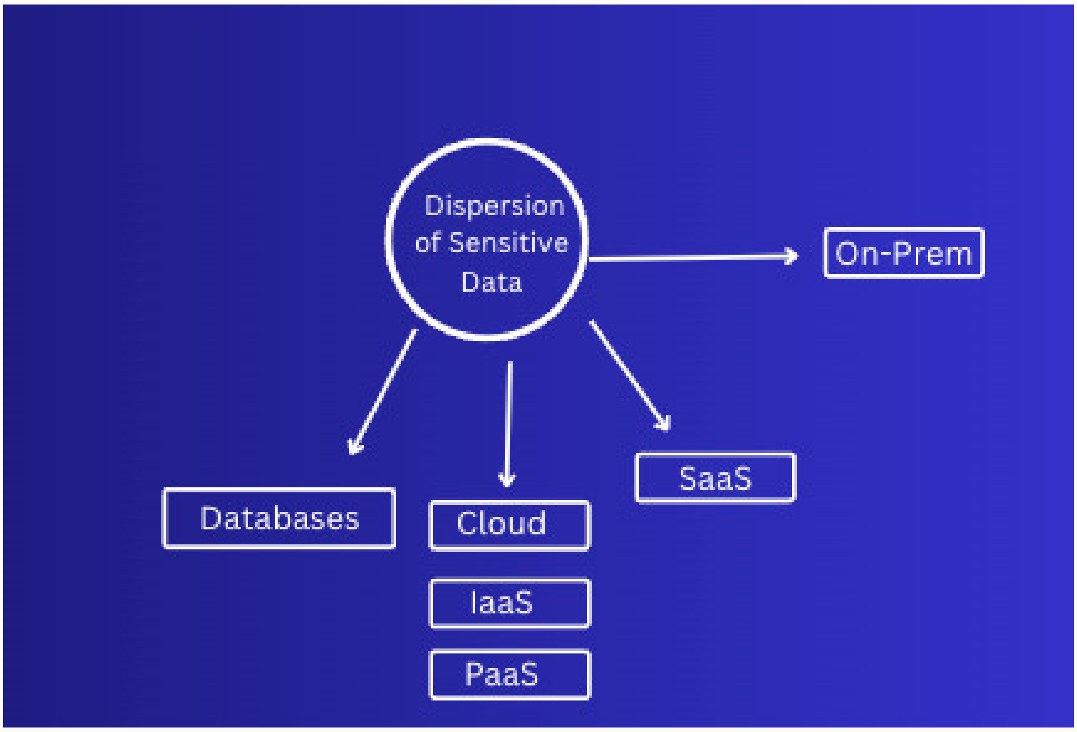
Defining Data and Its Context in Data Security

It's important to begin by defining data and its related context. Data can take various forms, including sensitive information, confidential data, cryptographic algorithms, and AI models. However, the focus of this report is on sensitive business data, which can manifest in various ways.

- **Structured data:** Data stored in a predefined format like databases, spreadsheets, and tables. Examples include customer records in CRM systems or employee data in HR systems.
- **Unstructured data sets:** Data that doesn't follow a predefined format, such as emails, documents, social media posts, images, videos, and chat logs. This can include internal communication threads, customer support tickets, and presentation files.

Where Is This Data Stored?

These datasets are usually stored across Databases, SaaS, Cloud and Web.

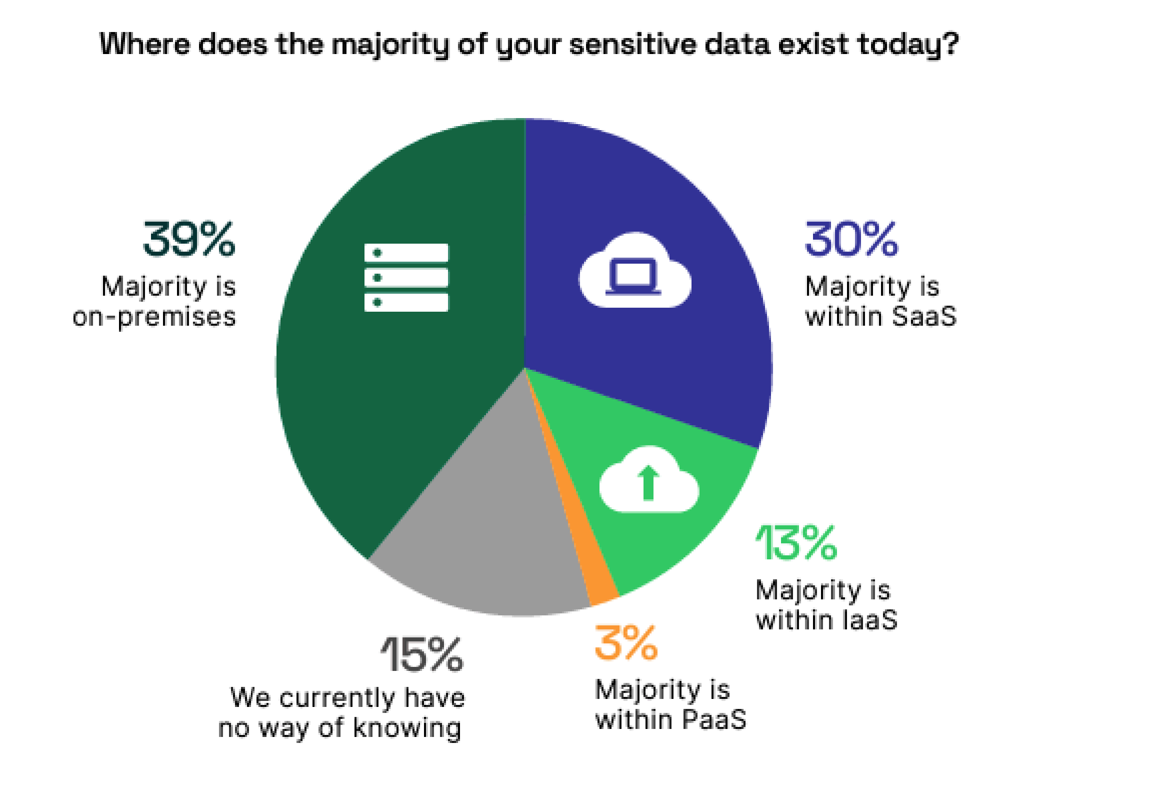


Across these locations, Data in cybersecurity takes on various forms, each presenting unique security challenges and requiring tailored protection measures.

- **Data at Rest** refers to information stored in static locations, such as servers or cloud storage. They are often secured through encryption, access controls, and data classification to prevent unauthorized access or tampering. In contrast,
- **Data in Transit**, moves across networks, is safeguarded with secure communication protocols like TLS/SSL and VPNs to prevent interception or tampering during transmission.
- **Data in Use**, actively processed by applications or users, requires runtime encryption and endpoint detection to protect against memory-resident attacks.

Visibility Remains A Challenge

However, the biggest challenge many enterprises face today is the dispersion of this data across multiple locations. Over 48% of information security professionals lack visibility into data within SaaS environments, highlighting the complexity of securing cloud-based data. 83% of respondents agree that a lack of data visibility weakens their security posture, highlighting its critical role in effective protection.



For those organizations that know where their data is located, over 39% of enterprises still report they have a large amount of data still within on-prem and 30% report SaaS locations showing the dispersion across enterprises.

Context is Key

Since organizations have different types of data, visibility must consider its relevant context. Organizations have different forms of data such as sensitive and regulated data, like PII, PHI, or financial records, to ensure compliance and avoid legal repercussions. But they also have other types of data, including behavioral and observational data, generated from network traffic or user activity, to support anomaly detection and threat intelligence, or for other business purposes. Therefore, enterprises need context around every data. They need to know:

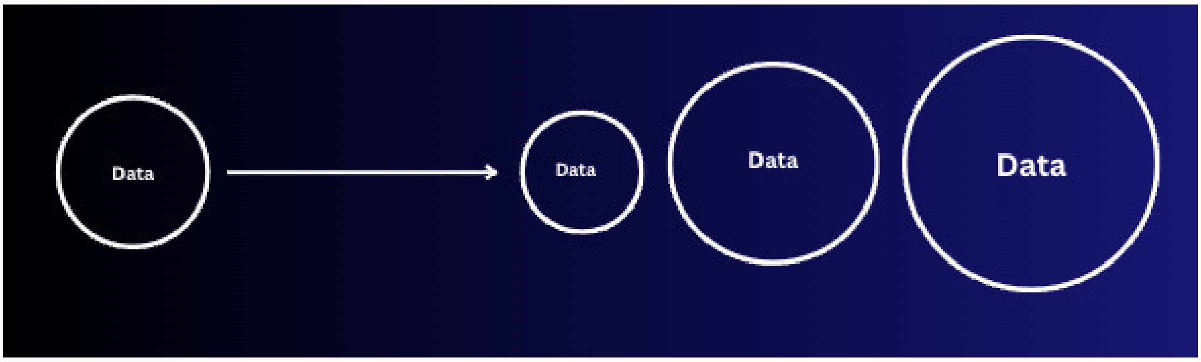
1. If data is identifiable (PII or not)
2. The role associated with the data and who should have appropriate access
3. Regionality (understanding the regulatory requirements in every region associated with that data)
4. The ability to derive metadata about data is crucial since every dataset is unique

All these factors highlight the need for better data security solutions.

Challenges Enterprises Face Today & The Driving Forces For Data Security

There are primarily seven key driving forces driving this focus on data security:

1. **Volume of Data Breaches:** Significant data breaches have impacted major companies, including 23andMe, AT&T, Ticketmaster, Dell, and American Express. A primary reason enterprises invest in DSPM solutions is to safeguard against potential data breaches (20%), followed by facilitating the deployment of GenAI technology (13%). Enterprises are focused on understanding where their sensitive data is stored and categorizing the contents of their data stores to take proactive measures against potential data leakage and loss. Volume of data breaches: There have been significant data breaches (23andMe, AT & T, Ticketmaster,Dell, American Express. One of the biggest reasons enterprises buy DSPM solutions is to protect them against potential data breaches (20%), followed by facilitating the deployment of GenAI technology (13%). Enterprises want to know where their sensitive data is stored and categorize what is inside of their data stores so they can take steps to avoid potential data leakage and loss.



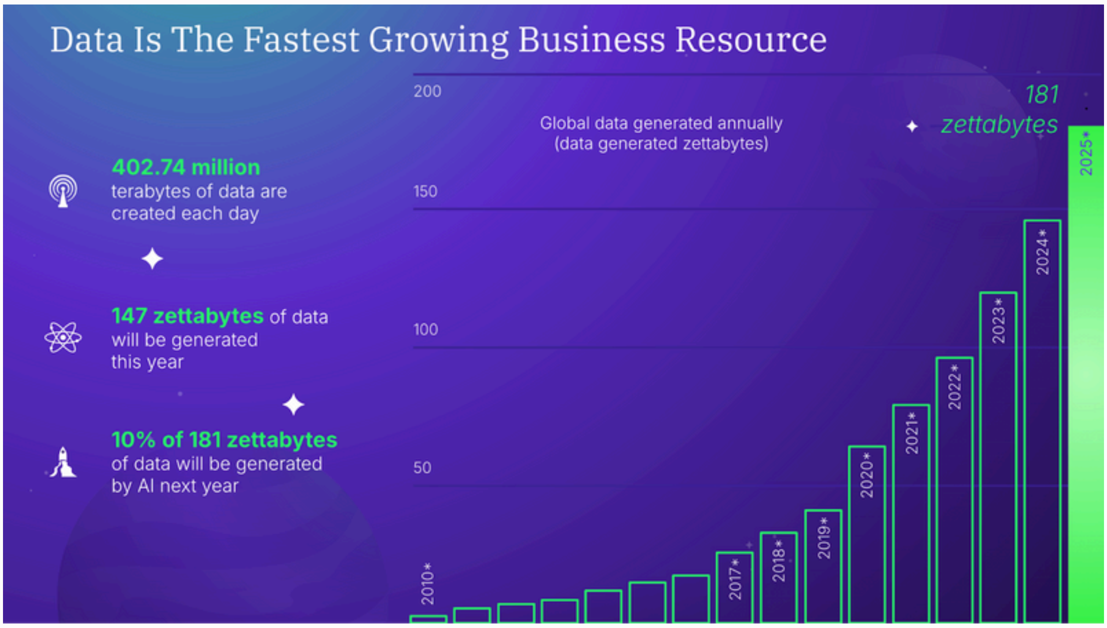
2. Cyber Resilience and Recovering from a Breach: Organizations face substantial financial losses due to downtime, management distraction, loss of intellectual property, and reputational damage. In fact, 53% of organizations reported experiencing a material loss of sensitive information in the past year alone. Notable breaches, such as those at Dell, American Express, Bank of America, and 23andMe, highlight the real costs associated with data security failures. According to IBM's annual "Cost of a Data Breach Report," the global average cost of a data breach reached $4.88 million in 2024, marking a 10% increase from the previous year.

3. Recovering from a Data Breach – Ransomware Fees: Over 80% of businesses impacted by data breaches opt to pay ransom fees, which totaled $1.1 BN in 2023Recovering from a data breach - Ransomware fees: Over 80% of impacted businesses pay ransom fees, which totaled $1.1bn in 2023.

4. **The Adoption of Gen AI Complicates Matters:** As businesses strive to remain competitive, data has become a crucial asset. The integration of data in GenAI has increased the need to locate and categorize data used by large language models to prevent sensitive information from being inadvertently included in these models and leaked. Model development often involves both open and closed-source models, along with extensive testing, resulting in some enterprises running over 100 different models. This complexity makes it difficult to identify security vulnerabilities across all models, leading to more than 40% of companies experiencing privacy or security issues tied to AI models. The rise of AI co-pilots is pushing more SaaS systems to demand access to enterprise data. While 64% of companies report feeling under pressure to adopt Generative AI, 84% view cybersecurity as the primary barrier to adoption.

5. **Data is the fastest-growing resource in the world.** The volume of data is rapidly expanding and leading to cloud data sprawl and associated risks: Data is one of the fastest-growing assets within organizations, but this growth often leads to data sprawl—where data is duplicated across multiple systems. This creates significant challenges in managing data, as disparate data sources cause a lack of visibility. With companies leveraging multiple cloud environments and SaaS platforms, managing data sprawl has become increasingly complex. As data grows, cloud resources become more vulnerable due to poor access controls, unsecured ports, and improper backup management. Notably, 99% of cloud identities are "excessively privileged," and 80% of data breaches involve data stored in the cloud, highlighting the risks associated with inadequate data management and oversight.

Approximately 402 million terabytes of data are created daily. About 147 zettabytes will be generated this year. 181 zettabytes of data will be generated in 2025. The average company had 100 SaaS Apps, 4-years ago and now it is over 3000+ SaaS Apps.

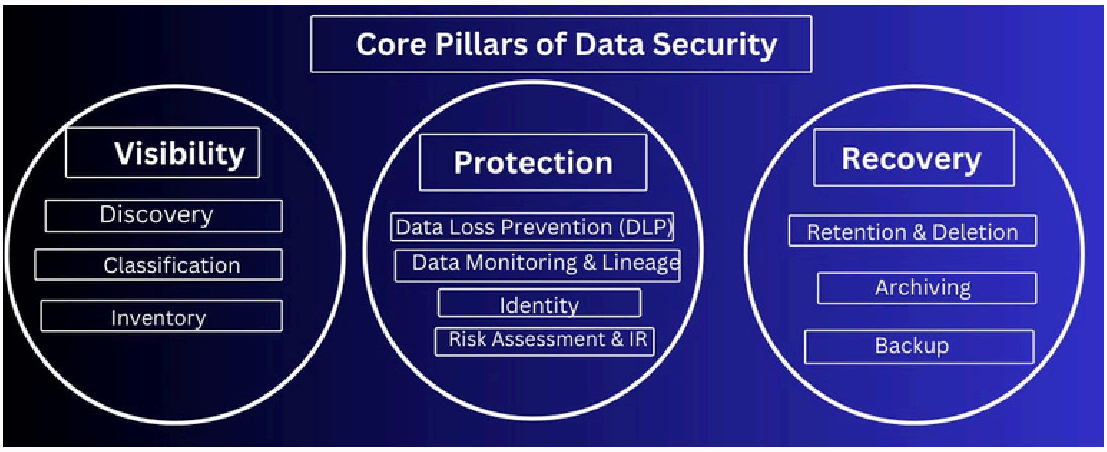


6. **Fines and Regulatory Compliance Penalties:** Compliance requirements such as HIPAA, GDPR, CCPA, DORA, LGPD, and PIPEDA are among the most prominent. With 137 countries enacting data protection legislation, the most notable being GDPR and CCPA, it is expected that privacy regulations will only continue to intensify, especially with the recent proposal of the American Privacy Rights Act.

7. **Compliance Frameworks and Sensitive Data Management:** Organizations are often unaware of the sensitive information they possess and where it resides, hindering their ability to safeguard it adequately. Only 4% of enterprises have dedicated storage for sensitive information.

Categories of Full Data Security Program

After a company develops a data security program, they should implement the following:



Data Discovery, Classification, and Inventory:

The foundational steps of data security consist of three critical components:

1. **Discovery:** This initial step involves systematically identifying and locating all data assets across an organization's environment.
2. **Classification:** This process involves categorizing and labeling data based on its sensitivity and importance, enabling the application of appropriate security controls.
3. **Data Inventory**: This involves maintaining a comprehensive catalog of both structured and unstructured data to ensure complete visibility and control.

Data Protection:

- **Data Protection and Lineage**: Modern Data Loss Prevention (DLP) systems integrate with Cloud Access Security Brokers (CASB) to create comprehensive protection. These systems monitor data movement, detect sensitive information, and prevent unauthorized access or exfiltration across all environments - whether data is at rest, in use, or in transit.
- **Encryption and Data Security:** Multiple layers of protection including data masking, encryption, tokenization, and hashing ensure sensitive information remains secure even if accessed. These technologies transform sensitive data into unreadable formats while maintaining functionality.
- **Identity and Access Management**: Robust access controls combine user rights management, behavior analytics, and continuous monitoring. The framework follows the principle of least privilege:

  Data Access Governance = Data Sources + Entitlements + Permissions + Enforcement Actions

  Risk Evaluation + Insight / Workflow = Remediation Actions
- **Security Monitoring and Response:** Comprehensive incident response through event triage, continuous monitoring, and integration with Security Information and Event Management (SIEM) systems ensures rapid detection and response to potential data security incidents.

Data Deletion, Backup and Recovery

- In an era where ransomware is one of the greatest attacks of our time, robust backup and recovery systems are more critical than ever. Organizations must implement comprehensive backup strategies. Regular testing of recovery procedures ensures business continuity and helps organizations quickly restore operations after potential incidents or data breaches.
- Enterprises should develop secure backup storage and implement redundant storage systems with proper access controls and encryption. Secondly, they should have a disposition protocol that implements secure deletion methods.
- With the rise of AI systems, data disposition has become increasingly complex, requiring careful consideration of regulatory compliance and ethical implications. When deleting data, organizations must ensure complete removal across all systems, including AI training datasets, cached versions, and backup copies.
- A foundation built around data discovery, protection and recovery sets the foundation for many other data use cases across the enterprise. Some of the top priorities for enterprises include securing data before it is used in AI models, particularly for Generative AI/LLM use cases. Other key areas of focus across the enterprise involve supporting Privacy, Governance, Risk & Compliance (GRC). This area emphasizes the proper handling, processing, and storage of personal and sensitive data to ensure both individual rights and privacy are protected, and compliance with regulatory frameworks is achieved.

Data Security Lifecycle Management for Enterprises

Data Security Posture Management (DSPM)

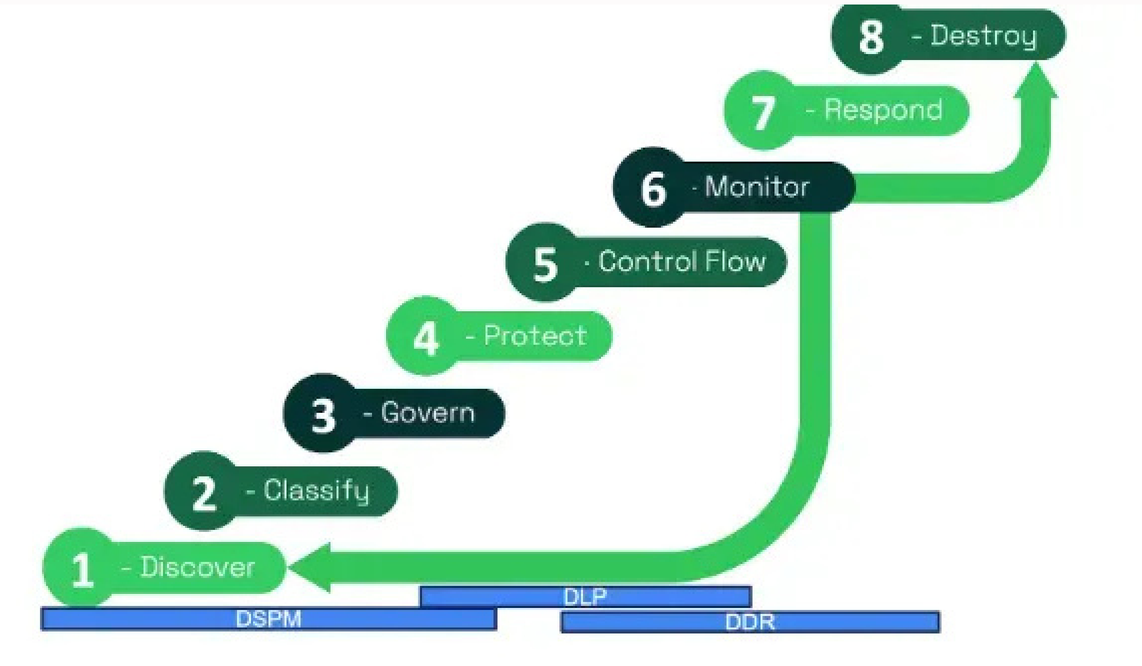
- Data Discovery
- Data Classification
- Data Governance

Data Protection

- Data Loss Prevention (DLP)
- Data encryption and masking
- Data monitoring and response

Data Recovery and Destruction

- Data backup and recovery
- Data deletion



Advanced Data Discovery

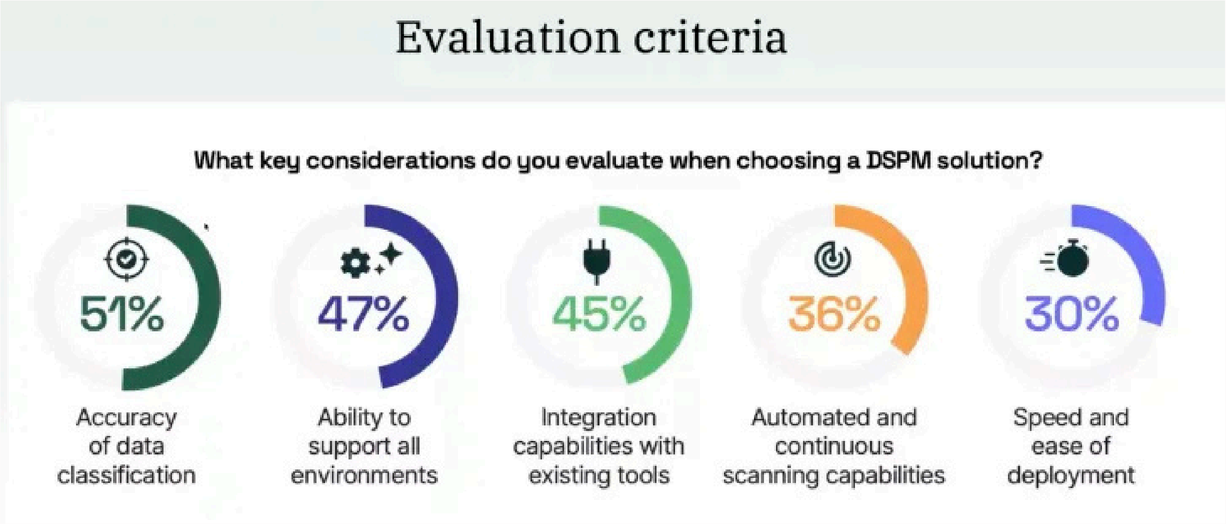
Sensitive data is often scattered across diverse repositories in today's complex enterprise environments, including cloud services, on-premises systems, and third-party applications. The first step is discovering where the data is and developing a centralized location for all enterprise data. Without discovery, efforts to secure data are undermined, making this step foundational for building a robust security posture.

1. **Identify All Datastores:** From structured databases to unstructured files in cloud storage, DSPMs uncover repositories that may otherwise go unnoticed. DSPM need to be able to discover data across:
    - Block stores
    - Object stores
    - Managed cloud databases
    - Managed cloud data warehouses
    - Self-hosted, embedded databases
    - Data stores in isolated private cloud environments
    - On-premises data stores (private data centers)

2. **Map Data Relationships:** By analyzing metadata and access patterns, DSPMs reveal how data flows between systems, enabling organizations to identify critical points of exposure. Comprehensive discovery enables organizations to identify sensitive data that might be improperly secured, such as data stored in misconfigured cloud buckets or shared folders with excessive permissions. Accurate data discovery facilitates adherence to regulatory requirements such as GDPR, CCPA, and HIPAA. By maintaining an up-to-date inventory of sensitive data, organizations can demonstrate compliance during audits and avoid costly penalties.

Data Classification

The ultimate objective of data classification for regulatory compliance is to ensure that an organization's sensitive data is accurately identified, categorized, and managed in alignment with applicable legal, regulatory, and industry standards. 51% of operators assess a DSPM vendor based on their classification efficacy.



Proper classification helps apply appropriate security controls to protect sensitive data, such as Personally Identifiable Information (PII), Protected Health Information (PHI), or financial records. It ensures compliance with specific requirements under laws like GDPR, HIPAA, CCPA, PCI DSS, and others.

- **Managing regulators:** It provides a clear and systematic method for regulators to verify that sensitive data is managed correctly, so, organizations can minimize risks of non-compliance, avoid costly penalties, and minimize impact from a potential breach.
- **Supporting Data Access and Retention Policies**: Data classification enables organizations to enforce role-based access controls, ensuring only authorized personnel can access sensitive data. It also helps manage data retention policies, ensuring compliance with regulations that dictate how long specific types of data must be retained or when it must be deleted.

Traditional Data Classification

Traditional data classification methods have focused on manually defining and categorizing data based on sensitivity. Key characteristics of these old methods include:

- **Long implementation timeline:** The old methods for managing data were flawed. According to data from Enterprise Strategy Group (ESG), after surveying 1000+ executives, they found that DSPM deployments typically take 4–6 months, following a 3-months timeline.

- **Rule-Based Systems:** Relied on static patterns, such as keywords, regex (regular expressions), or predefined rules. For instance, regex could flag “Jordan” as sensitive but could not discern whether it referred to a person, a country, or a brand. This context-blind approach often resulted in false positives, undermining their reliability in complex data environments.

- **Manual Data Labeling:** Business teams were assumed to understand their data and performed manual label sensitivity of data. This model also proved untenable in environments with rampant data sprawl and dynamic workflows, where sensitive data could reside across multiple siloed ecosystems.

- **Static Detection Algorithms:** Static detection methods, such as Exact Data Matching (EDM) and file fingerprinting, were decent advancements over manual labeling (focused on identifying exact matches to predefined data patterns.). While effective for structured data, they were inefficient and costly for unstructured or semi-structured data.

The challenges with these solutions was that they had high false positive rates and lack of contextual understanding. They were resource-intensive and time-consuming processes, unsuitable for modern, large-scale environments. They also could not adapt to new data types or evolving business needs. Additionally, traditional methods were resource-intensive, often requiring weeks or months to complete scans in large repositories, leaving critical gaps in data security.

Modern Approaches to Data Classification with the Advent of AI and LLMs

AI-powered models are inherently adaptive and capable of contextual understanding. LLMs, trained on diverse datasets, can analyze complex, unstructured data formats such as emails, reports, and hybrid semi-structured files. This adaptability enables detection of sensitive data types that traditional systems often miss.

Let’s take the example of “Jordan”

- Jordan is a country
- Jordan is a popular brand
- Jordan is a shot brand

The best way to improve classification is by leveraging contextual data, allowing AI to determine the context in which this data is used.

AI has its own limitations, such as measurement difficulties, lack of specificity, hallucinations (false information risks), and control and consistency issues. Additionally, costs remain high even when using OpenAI AIs. This means classification requires more sophistication. Companies like Cyera have enhanced modern methods leveraging RegEx, NLP, Statistical validation and AI. Cyera uses a classification engine with a high recall model (cleaning the garbage) + a high precision model + context, which has proven to achieve over 90% accuracy in their classification, compared to competitors.

Contextual and Metadata:

- AI models assess the surrounding metadata and usage context of data, enabling nuanced classifications. For example, distinguishing between a customer's phone number and an employee's phone number allows for tailored security protocols. Modern systems enrich classification outcomes with metadata - such as geographic location, compliance frameworks, and data roles, enabling granular policy enforcement.

Objectives for customers:

- The best AI-powered classification engines continuously learn and adapt to unique data environments. Moving beyond content-level detection, file-level classification considers the entire file's characteristics, enabling the identification of sensitive documents like financial reports or intellectual property
- By reducing false positives and scaling security efforts, AI-based systems enhance the understanding of data, minimize manual efforts, reduce security gaps, and offer insights and recommendations.

Data Access Governance

Leveraging core identity security protocols and Data Access Governance (DAG) are critical components of modern data security that enables organizations to implement least privilege access at scale while maintaining comprehensive visibility into data activity. At its core, DAG combines several key elements: data sources, entitlements, permissions, and enforcement actions, all working together to create a robust access control framework. The foundation of DAG lies in its access control policies, which determine both who can access specific datasets and under what circumstances.

Data & Identity Security

Identity and data have always been interconnected, like two sides of the same coin. Yet for years, organizations have primarily focused on the identity side. While identity-driven approaches like least privilege access and Zero Trust frameworks are effective, they often fall short without the visibility and context of the data they aim to protect. For example, we often know who has access to what, but lack insights into how that access is used or if it is even necessary. This gap leaves sensitive data—now the fastest-growing attack surface—exposed, particularly in areas like insider threats, compliance risks, and third-party access.

By integrating identity context with data insights, organizations can achieve a unified understanding of who and what interacts with sensitive information, uncovering previously unanswered questions. This symbiosis enables granular access control, stronger risk management, and enriched security postures. For instance, understanding data flow—from application to system to user—provides visibility into data sprawl, potential misuse, and weaknesses in protection mechanisms. Similarly, tracking non-human identities and external access reveals critical risks tied to third-party integrations. Data Security will both need to build new solutions, but have strong integrations with identity security vendors. For effective implementation, DAG must integrate with existing security infrastructure, including Single Sign-On (SSO) and Identity Providers (IdPs).

Data Protection

There are three components to data protection:

- Data Loss Prevention (DLP)
- Data encryption and masking
- Data monitoring and response

Data Loss Prevention (DLP)

The core objective of DLP solutions is to prevent unauthorized data transfers and leaks by ensuring that only authorized systems or users can transfer sensitive data. They were designed to monitor and block data movement that violates policies. DLPs are installed across endpoints, cloud, email, SaaS (CASB). Companies also have some form of DLP across Web, AI models and at the API.

Traditional DLP

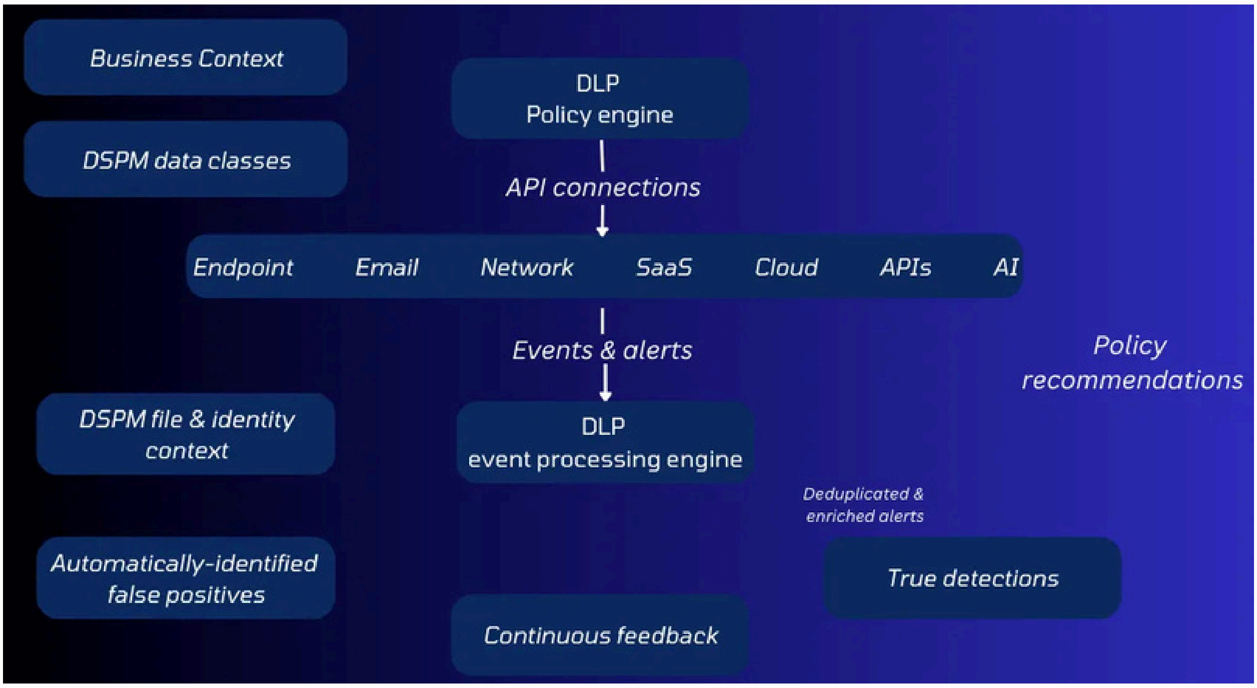
Symantec and McAfee built the first generation of DLPs. These legacy DLP solutions relied on rigid, rule-based and heavy agent-based systems that struggled to adapt to modern data environments. Their manual rule creation led to significant challenges such as:

- Poor accuracy and high false positives: These systems generated up to 90% false positives because of the wrong classification, overwhelming incident response teams with many unnecessary alerts.
- Limited intelligence—lack of contextual awareness meant systems couldn't distinguish between legitimate business activities and actual threats.
- Complex deployment—agent-based models required years to implement, with significant operational overhead for enforcing policies. Further, there was a fragmentation and long timeline for detection across endpoints, email, and networks, creating inconsistent policies and security gaps.

Ultimately, traditional DLP solutions lacked the flexibility, accuracy, and efficiency needed to protect sensitive data in today's dynamic environments. These shortcomings made the need for more sophisticated and integrated solutions clear

Modern DLP

Modern AI-driven DLP solutions leverage agentless technology, combining advanced AI, real-time enforcement, and seamless integration to create a dynamic, adaptive framework. These solutions effectively protect sensitive data across all environments, meeting the demands of today's complex data landscape. Below is a sample of a modern DLP based on work with Trail's DLP solution (now acquired by Cyera).



Trail DLP

- **DSPM Classification Engine:** Traditional DLP solutions lacked context from DSPM systems, but modern solutions, such as those used by Cyera, integrate DSPM with DLP to inform detection capabilities, significantly reducing false positives. Cyera's agentless DLP enables real-time detection with proactive alerts. AI-powered DLP can analyze data movement in real time, stopping unauthorized transfers before they occur.

- **Agentless Models:** Modern DLP solutions can be deployed quickly using APIs to integrate with existing tools, eliminating the need for extensive agent installations.

- **Anomaly Detection**: Behavioral analytics detect unusual user or system activities, providing early warnings for potential breaches.

- **Self-Tuning Policies:** AI continuously refines DLP policies based on real-world data, reducing false positives and improving accuracy over time through self-learning and feedback loops.

The Future of AI-powered DLP lies within the context of DSPM solutions:

Cyera acquiring Trail Security is the perfect illustration of this dynamic at play. As discussed, by combining DSPM's labeled data with a GenAI-powered DLP engine, DLP solutions today have become far more capable than they were 20 years ago. All of this is powered by more data, leading to more accurate detections, because the use of more discovered and classified data improves the accuracy of data in-motion.

Leveraging DSPM/AI to build DLP allows for easier creation of detection rules and logic for DLP engines. Lastly, all of this improves the visibility of data at-rest, in-motion, its lineage as well as identifies all points of flow. As a result now, organizations are better able to use much more enhanced DLP to prevent insider threats like IP, customer or accidental mishandling. They can better protect against AI data use and unintentional sensitive data exposure by LLMs and prevent oversharing with 3rd parties.

Data Encryption and Masking

Data encryption is another crucial aspect of data security. It involves using various techniques to safeguard sensitive information from unauthorized access, both during transmission and storage. Encryption ensures that even if data is intercepted or accessed without proper authorization, it remains unreadable without the correct decryption key. Some enterprises will leverage some form of hashing of their data. In this piece, we'll focus on these two methods:

1. **Data Encryption:** Encryption is a key method used by enterprises. It obfuscates data through techniques like encryption, tokenization, and masking, while also managing encryption keys. A comprehensive encryption strategy can help companies follow regulations and serve as a last line of defense if data is breached. Sensitive data should always be subject to some form of encryption, but traditional encryption can make data difficult to use. Companies might consider technologies like tokenization (a key tool used by **credit card networks**), format preserving encryption, and homomorphic encryption (which could power machine learning on encrypted data). For example, Skyflow is building a new piece of this architecture, the data privacy vault - a centralized point of control for sensitive data– to ensure that data doesn't proliferate across a company's systems
2. **Data Masking:** This is used to conceal sensitive data from unauthorized users, typically through dynamic data masking based on policies such as ABAC. It is increasingly used as a core control for enabling data provisioning and access and in solutions that speak to a data audience (Chief data office and analytics officer).

Monitor and Respond

The third key component to data protection is ensuring that companies have good monitoring controls in the case of a detection, so they can respond effectively to incidents. 43% of professionals prioritize real-time data monitoring and alerting of data events based on data from cybersecurity insiders when deploying data security solutions.
Since we know that attackers can bypass security controls, data security platforms must include robust monitoring capabilities to detect risky user behaviors and verify proper data encryption and access controls. Some of the core protocols that key solutions have include:

- **Alert management:** Data security platforms need to integrate with SIEMs, ticketing systems (JIRA, ServiceNow), Email solutions (Gmail, Outlook) and Messaging tools (Slack, Microsoft Teams) to ensure rapid notification when threats have been detected in real-time.
- **Monitoring capabilities**: These include file integrity monitoring, third-party risk monitoring (TPRM), and database tampering detection. These solutions need data mapping and lineage capabilities to track data movement throughout the system.
- **User activity monitoring and behavioral analytics**: These should include capabilities for real-time threat identification to detect potential threats.

Data security platforms should be able to detect data compromises and support immediate forensic incident response. In the event of a data breach, solutions must provide comprehensive data points, through monitoring and mapping features, to enable thorough investigations.

Additionally, organizations should maintain a program that keeps their crown jewels backed up and ready for restoration. These solutions must support audit requirements. In the event of a ransomware attack where attackers gain access to information, organizations should be able to restore their crown jewels swiftly to resume business operations.

Data Recovery and Destruction

The final phase of an effective security program is ensuring that companies have strong data backup, recovery and removal capabilities of data that should not be used. Organizations often struggle with excessive data bloat and it primarily comes in three ways:

1. Data that is susceptible to malicious attackers and insider threats
2. Data that leaves an enterprise susceptible to data privacy fines
3. Data that is redundant, leading to excess storage costs

Hence, data security platforms should be able to support data sanitization and data destruction to ensure that sensitive or obsolete data is irreversibly deleted, leaving no residual information that can be exploited. Data clean-up addresses the accumulation of redundant, obsolete, or trivial (ROT) data, which can unnecessarily increase an organization’s attack surface. We’ve seen new compliance mandates like GDPR, CCPA, and HIPAA require enterprises to securely delete personal or sensitive data when it's no longer necessary. Failing to comply can lead to hefty fines. Hence, DSPM solutions should be able to allow data sanitization that ensures compliance with "right to be forgotten" and other data erasure requirements.

In addition, DSPM tools should be able to identify data repositories containing expired, redundant or sensitive information, enabling automated alerts for clean-up or destruction. Also, sanitizing and cleaning up old or unused data can lower storage costs and compute overheads, particularly in cloud environments.

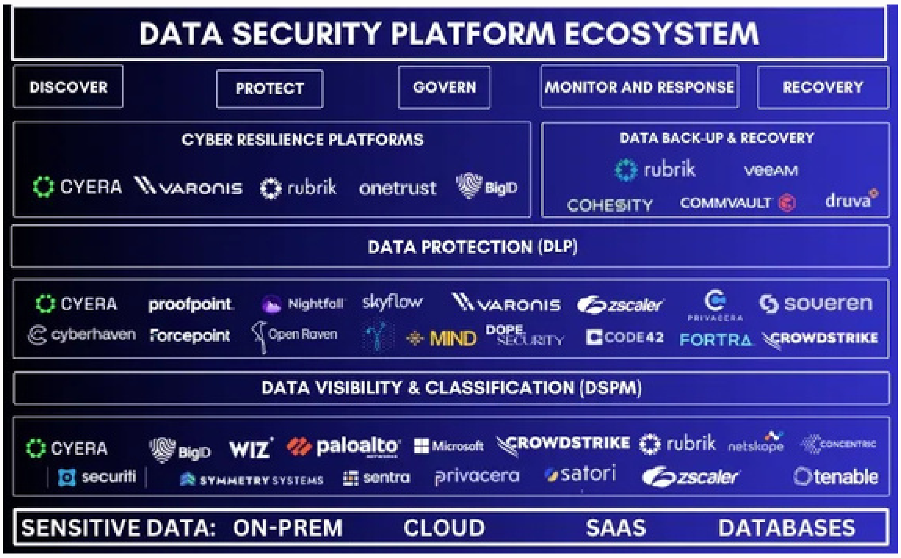
Lastly, organizations must have robust systems to locate and retrieve personal data across their environments to respond to DSARs (Data Subject Access Request) where an individual requests the deletion (or a copy) of their personal data under specific circumstances (e.g., when it's no longer necessary for processing). With new regulations, these have to be done promptly (usually within 30 days under GDPR).

Data Backup

Data backup and recovery are essential components of this strategy. DSPM platforms need to integrate and work alongside data backup companies to ensure data is properly backed up. The objective is to provide critical safeguards against data loss, breaches, and operational disruptions. When data breaches occur, companies need a way to resume operations. Backup systems create immutable copies of data, ensuring that ransomware attacks cannot encrypt or delete backups. Companies like Cohesity and Rubrik use immutable storage and snapshot chaining to protect against unauthorized changes to backup files. These solutions help with Business Continuity and Disaster Recovery (BCDR) in the case of hardware failures, breaches, or disasters. These backup solutions provide audit trails, retention policies, and data encryption to meet compliance requirements.

Data Security Platform Competitive Landscape

The data security market ecosystem has rapidly grown to many vendors over the past four years across the different categories outlined above. Some vendors have acquired their way into the market, while others have built their solutions from ground-up to compete in the market. The larger security vendors have acquired many of the DSPM vendors that were built over the last few years. According to the Altitude Cyber dataset, we have seen over $10B worth of investment and 500+ funding activities across the data security and adjacent market category.



DATA SECURITY PLATFORM ECOSYSTEM

DISCOVER | PROTECT | GOVERN | MONITOR AND RESPONSE | RECOVERY

CYBER RESILIENCE PLATFORMS

DATA BACK-UP & RECOVERY

DATA PROTECTION (DLP)

DATA VISIBILITY & CLASSIFICATION (DSPM)

SENSITIVE DATA: ON-PREM CLOUD SAAS DATABASES

The competitive landscape can broadly be divided into the following players, with in-depth coverage across the following infrastructure areas (important to mention that just because a vendor has strengths in an area doesn't mean they don't cover other areas):

- **On-prem:** Varonis, Rubrik (Laminar), BigID
- **Cloud:** Cyera, Wiz, Prisma Cloud
- **SaaS:** Microsoft Purview (M365), Netskope, Salesforce (Own)
- **Data In-Transit**: Crowdstrike, Zscaler, Soveren
- **Native solutions**: Satori, Privacera, Sentra, Concentric AI, Securiti are a few of the standalone vendors in the market today

Once again, there are many vendors outlined in the market map. It's important to clarify that some of these vendors have expanded their offerings to multiple areas with data.

Many of these vendors have security controls across DSPM (Discovery & Classification), DLP (Protecting how the data is accessed, what the data is and its encryption in-transit), and finally, the backup of the data.

After 10+ companies have been acquired by large platforms, Cyera has emerged as the most funded, fastest-growing, and standalone holistic solution across most of the areas outlined above. Hence, the rest of the report focuses on explaining most of the theoretical concepts using an existing platform, as a case study.

Cyera

Founded in 2021 by Yotam Segev and Tamar Bar-Ilan, Cyera has quickly emerged as one of the few standalone data security platforms designed to provide comprehensive visibility, control, and protection of data assets, regardless of their location. Based on discussions with CISOs and leading companies, Cyera has experienced the most significant growth and traction amongst all data security platforms on the market. Here are some highlights:

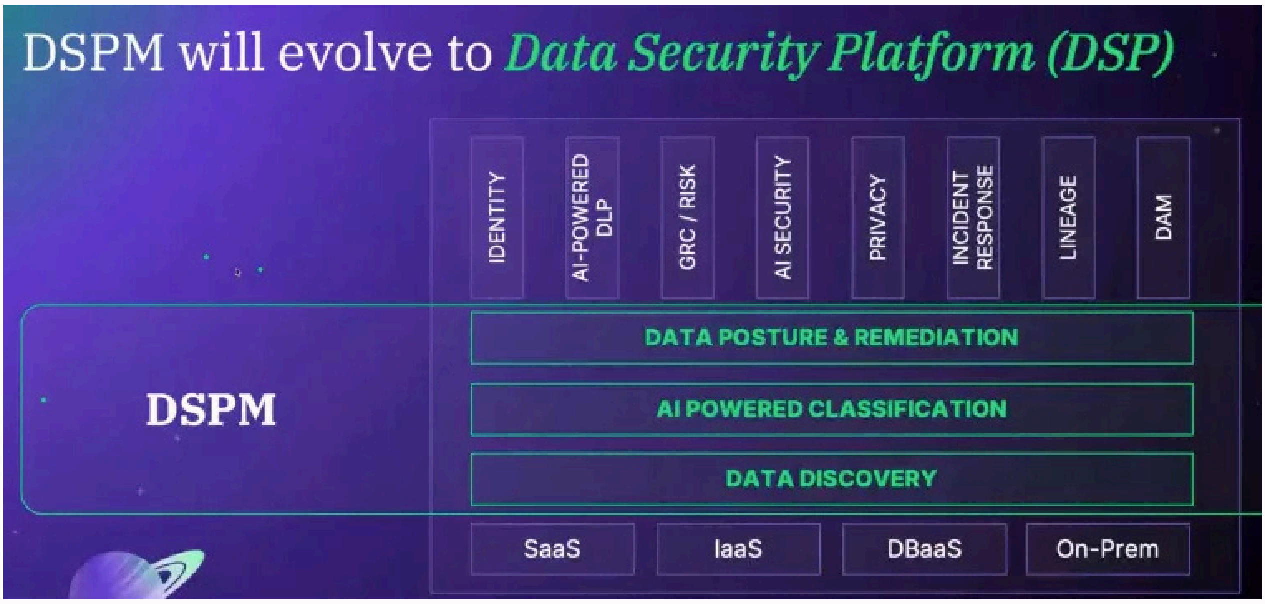
1. They have the highest valuation within the data security market for companies of their size.
2. They have raised over $700+ million in funding.
3. They have 200+ customers globally.
4. They are growing revenues by 4.6x.

The company's core value proposition centers around enabling enterprises to secure their data across the entire lifecycle — from discovery to destruction.

Cyera’s Philosophy: A Rubik’s Cube Model of Building An Intelligent Data Security Brain

Cyera is building a solution which, first of all, integrates with all your existing security tools such as EDRs, network security, SaaS security and many others to discover data in all its disparate locations. Once they have given you the full visibility, they then apply their classification engine to understand the context for data across the enterprise. Then, they apply their AI engine to derive context, risk, and enforce control to build a “cohesive brain” across the enterprise stack. By leveraging these foundational principles of a core understanding of the data, they are able to use this “data security” platform to pass along insights to other solutions.

Cyera’s Platform



DSPM will evolve to *Data Security Platform (DSP)*

IDENTITY
AI-POWERED DLP
GRC / RISK
AI SECURITY
PRIVACY
INCIDENT RESPONSE
LINEAGE
DAM

DSPM

DATA POSTURE & REMEDIATION
AI POWERED CLASSIFICATION
DATA DISCOVERY

SaaS
IaaS
DBaaS
On-Prem

Cyera is building a solution that covers the entire data security lifecycle from discovery to destroy.

1. **Discovery and Classification**: Cyera’s solution has proven itself to accurately discover & classify data in real-time across the enterprise. It can govern data access & use by identifying contextual risks such as excessive access permissions or misconfigured storage buckets. They have built a solution that cuts across discovery and classification as the foundation. They have identity-based data access controls, allowing organizations to monitor how employees interact with specific datasets.
2. **Discovery, Detection and Response (DDR) powered by DLP:** Cyera’s solution builds upon all the security controls I outlined earlier, allowing organizations to monitor who has access to which data and how it is being leveraged internally. The recent Trail Security acquisition enhances their protection capabilities, which will be further developed in the coming year. Cyera offers an agentless solution that provides a unified view across fragmented DLP systems. It uses AI to better inform DLP policies and improve existing solutions. With real-time detection and orchestration, Cyera extends DLP coverage to more endpoints and applications, providing detailed lineage and data flow insights. Their platform has proven to help enterprises respond effectively to data breaches, offering breach scope audits in case of regulatory involvement. Additionally, the platform enhances remediation workflows, identifying vulnerabilities and assessing risks.

3. **Identity-Centric Data Security:** Cyera's approach to identity security redefines data protection by addressing the intertwined challenges of data and identity management. By unifying contextual data insights with identity attributes, Cyera gives organizations visibility and control over sensitive data. From my perspective, Cyera has taken a fresh and much-needed approach to solving one of the biggest challenges in data security: connecting identity with data in a meaningful way. For years, we've leaned heavily on identity-driven decisions—like least privilege access—but often lacked the full context of the data itself. Cyera flips the script by unifying data visibility with identity context, making it possible to answer those critical, unanswered questions about who truly has access to sensitive information, how it's being used, and where the risks lie. What stands out to me is how Cyera goes beyond just solving for compliance or insider threats. It's about enabling businesses to truly understand the flow of their data—across applications, systems, and users—in a way that enhances security while driving innovation. Whether it's identifying over-permissioned access, ensuring third-party access is secure, or tackling sprawling data environments, Cyera's solution provides clarity where most tools fall short. This is the kind of holistic, data-first thinking that the future of security demands

Cyera's Competitive Advantage

There is no doubt that Cyera competes with established players like Varonis, and BigID, as well as emerging startups such as Laminar and Dig Security. Cyera's ability to combine data discovery, classification, and risk mitigation into a single platform gives it a competitive edge. Based on my extensive research, Cyera demonstrates several key competitive advantages over other players in the market:

The Data Security Journey
Discover
Classify
Govern
Protect
Control
Monitor
Respond
Destroy
CYERA

Superior data classification

Cyera's superior data classification capabilities are evident in their impressive 85-99% accuracy rate for sensitive healthcare data classification, and their overall 92% accuracy in sensitive data identification - a significant improvement over the industry standard of 50%. The company leverages advanced techniques, including the Luhn check algorithm for credit card data validation, to achieve this high level of accuracy. From a technical standpoint, Cyera stands out with its agentless solution for on-premises environments, surpassing traditional tools like Stealthbits, while also offering more robust Data Security Posture Management (DSPM) technology compared to competitors like Wiz.

Cyera just reached a milestone of successfully classifying 1 exabyte (1000 petabytes) of Snowflake data and achieving over 95% precision rate in data classification. For context - this is the count for every sand element across the world. They were able to discover 1 trillion of sensitive records at risk.

A big component of how they achieved this goal is by using Cyera's advanced AI system that auto-learns new classifications and provides context (identifiability, security, role, geolocation). They combine natural language processing, machine learning, statistical validation + regex to make this happen, so it has compounding and reinforcement benefits that gives them a significant advantage over competitors over time.

The models are pre-trained on huge amounts of data to create robust, out of the box data classifiers to identify common data types - i.e. credit cards, SSNs, etc.

More importantly, the ability for Cyera's models to auto-learn new classifications that are unique to each customer is key. This is done from customer-specific data during production runtime, i.e. Employee IDs, product CKUs, lot numbers, claim numbers etc. are part of the auto-learning process. As the model adapts, the accuracy gets better and better. Between 40-70% of data Cyera classifies, is unique to the customer, which is crucial.

They continuously train their classification engine. The models classify data by analysts database metadata, file contents, and other contextual information. They ensure that only high-precision classifications, supported by a large amount of training data, are presented within the platform, minimizing false positives. A similar notion is how GraphRAG systems are used to enhance search through the use of context. Cyera applies the same to DSPM.

Speed of Classification

Another key element to highlight about Cyera's solution is the speed at which they are able to classify large amounts of data. In the case study above, they were able to classify 1000 petabytes of data.
A customer with 28.5TB of Snowflake data deployed Cyera at 3:00 PM to scan their environment for sensitive data. By 8:00 AM the next day, Cyera had completely scanned the environment, identified 1.6 billion at-risk sensitive records, and mapped them to global compliance frameworks like GDPR, HIPAA, and PCI DSS.

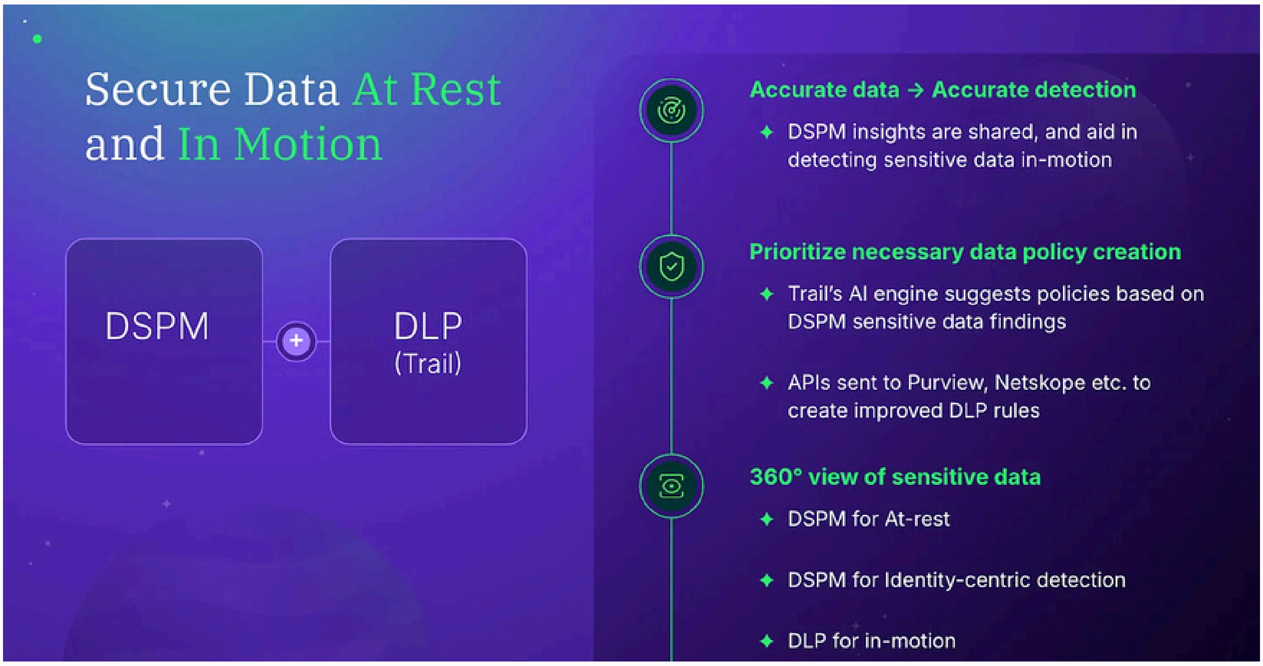
This process, which previously could have taken months, was completed by Cyera in less than a day, allowing security teams to quickly gain actionable insights and focus on risk reduction and innovation rather than data classification. The rapid growth of data, in the example of Snowflake, has resulted in an exponential increase in sensitive cloud records, making this type of rapid and efficient data scanning increasingly essential. Cybersecurity has lacked the tools to manage such large amounts of sensitive data, leaving many organizations exposed to higher risks of data breaches, compliance issues, and exposure.

Building A Data Brain

In terms of market position, Cyera has established itself as a leader through its focused approach, emphasizing speed and effectiveness in its solutions. The company has consistently outperformed competitors like Laminar (Rubrik), whose progress has been hindered by acquisition-related delays. Additionally, Cyera's broader coverage in data security, compared to Wiz's limited scope, has helped them secure deals in the majority of competitive sales cycles, further solidifying their position in the market.

Cyera is building a data security platform on the pillar of Data posture management (DSPM) and Data Loss Prevention (DLP), all powered by Gen-AI, which drives all other security programs within the enterprise.

Key piece here is Cyera positioning themselves as the “data brain” that passes along key insights to other key technologies, such as Security Service Edge Services (SSE), CSPM (i.e. Wiz) SIEM, Backup & recovery, Endpoint Security, Snowflake, DevOp Security (Armor Code), and even Microsoft Purview etc.



Secure Data At Rest and In Motion

DSPM + DLP (Trail)

Accurate data → Accurate detection

- DSPM insights are shared, and aid in detecting sensitive data in-motion

Prioritize necessary data policy creation

- Trail’s AI engine suggests policies based on DSPM sensitive data findings
- APIs sent to Purview, Netskope etc. to create improved DLP rules

360° view of sensitive data

- DSPM for At-rest
- DSPM for Identity-centric detection
- DLP for in-motion

No one else is combining this level of depth of DSPM - Discovery, Classification, and overall posture - with Data Loss Prevention. Cyera's DLP (acquired Trail) uses the DSPM insights and AI to generate recommendations for existing DLP policies (from Purview, Zscaler, and others) to make them more accurate. It will also block sensitive data from flowing out as well.

The foundation pillar is DSPM (across IaaS, SaaS, DBaaS and on-premises). Then mapping data to identity (which is critical for zero trust data access), Assessment Services (Data Risk as well as Breach Readiness), leveraging their combination of DSPM + AI-Powered Data Loss Prevention. Later down the road, Cyera plans on building out a product module focused on risk - addressing privacy concerns for organizations.

Concluding Thoughts / Platform Opportunities

- In conclusion, the data security landscape is undergoing a seismic shift. The exponential growth of data, coupled with the rise of generative AI and an increasingly stringent regulatory environment, has made traditional network-centric security obsolete. Data itself is now the perimeter, demanding a data-centric security strategy.
- This report has illuminated the multifaceted challenges organizations face: data breaches, cyber resilience, the complexities of GenAI, data sprawl, compliance penalties, and the dynamic nature of data. With an increasing number of data privacy laws and compliance requirements, organizations are struggling to maintain compliance and avoid costly penalties. These challenges necessitate a comprehensive data security program that encompasses discovery, classification, governance, and protection, as well as robust monitoring, response, and destruction capabilities.
- Advanced technologies, such as AI-powered classification and agentless DLP, are no longer luxuries but necessities for effective data security. As data continues to grow and evolve, organizations will prioritize robust, holistic data security measures and platforms best positioned to protect their most valuable asset and importantly, ensure long-term resilience. The future of security is at the data source — both for business' competitive advantage as well protecting companies against cyber attacks