Securing the Data Landscape with DSPM and DDR

Understanding how data security posture management (DSPM) with data detection and response (DDR) reduce the risk of data exfiltration

Why DSPM? The Shift to Data-Centric Security

The global datasphere is projected to grow by more than 50%—from 120 zettabytes (ZB) in 2023 to 181 ZB by 2025.[1] It's no wonder the idea of data breaches is keeping CISOs up at night. The combination of digital transformation processes, an increased appetite for data and analytics, and the proliferation of cloud datastores means enterprises are storing more sensitive data than they can effectively monitor or control.

Driving the vast majority of breaches—94.6%, to be exact—remains financial motives.[2] Data have always been a target for hackers, but the overall cost of recovering from breaches, specifically ransomware incidents that result in costs, has doubled in a two-year span, according to the 2023 Data Breach Investigations Report (DBIR).[3]

Enterprises recognize that sensitive data puts them at risk, and current solutions lag behind the adoption of cloud data infrastructure. This has led to the emergence of data security posture management (DSPM).

DSPM addresses core challenges stemming from sensitive data stored across various cloud repositories. Designed for environments with multiple clouds and numerous services, DSPM equips organizations with practical tools to discover and secure sensitive data.

[1] *Data Created Worldwide 2010-2025*, Petroc Taylor. Statista. Statista. August 22, 2023.

[2] *2023 Data Breach Investigations Report*, Verizon Business. 2023.

[3] The 2023 Data Breach Investigations Report (DBIR) highlights that, based on breach impact data from their partner, the FBI Internet Crime Complaint Center (IC3), the cost of incidents (that incurred costs) has doubled since they reviewed data in the 2021 DBIR.

Understanding Data Security Posture Management (DSPM)

Data security posture management (DSPM) consists of various practices and technologies designed to assess, monitor, and minimize risks related to data residing in cloud datastores, particularly across multicloud environments. Emphasizing the protection of sensitive information, it examines the context and content of the data and prioritizes personally identifiable information (PII), medical records, and other critical information.

How Does DSPM Work?

While DSPM gains traction and recognition in the industry, it remains an emerging technology. Although we see ambiguity in how vendors and analysts describe DSPMs, most descriptions include the capabilities we outline in this guide.

Data Discovery: Identifying Where Sensitive Data Is Stored in Cloud Environments

DSPM tools provide visibility into your cloud data inventory—the various services where sensitive data are stored across IaaS, PaaS, and DBaaS deployments. This could include managed cloud warehouses (Amazon Redshift, Google BigQuery, or Snowflake), unmanaged or semi-managed databases running on virtual machines, and object storage (Amazon S3, Google Cloud Storage, or Azure Blob).

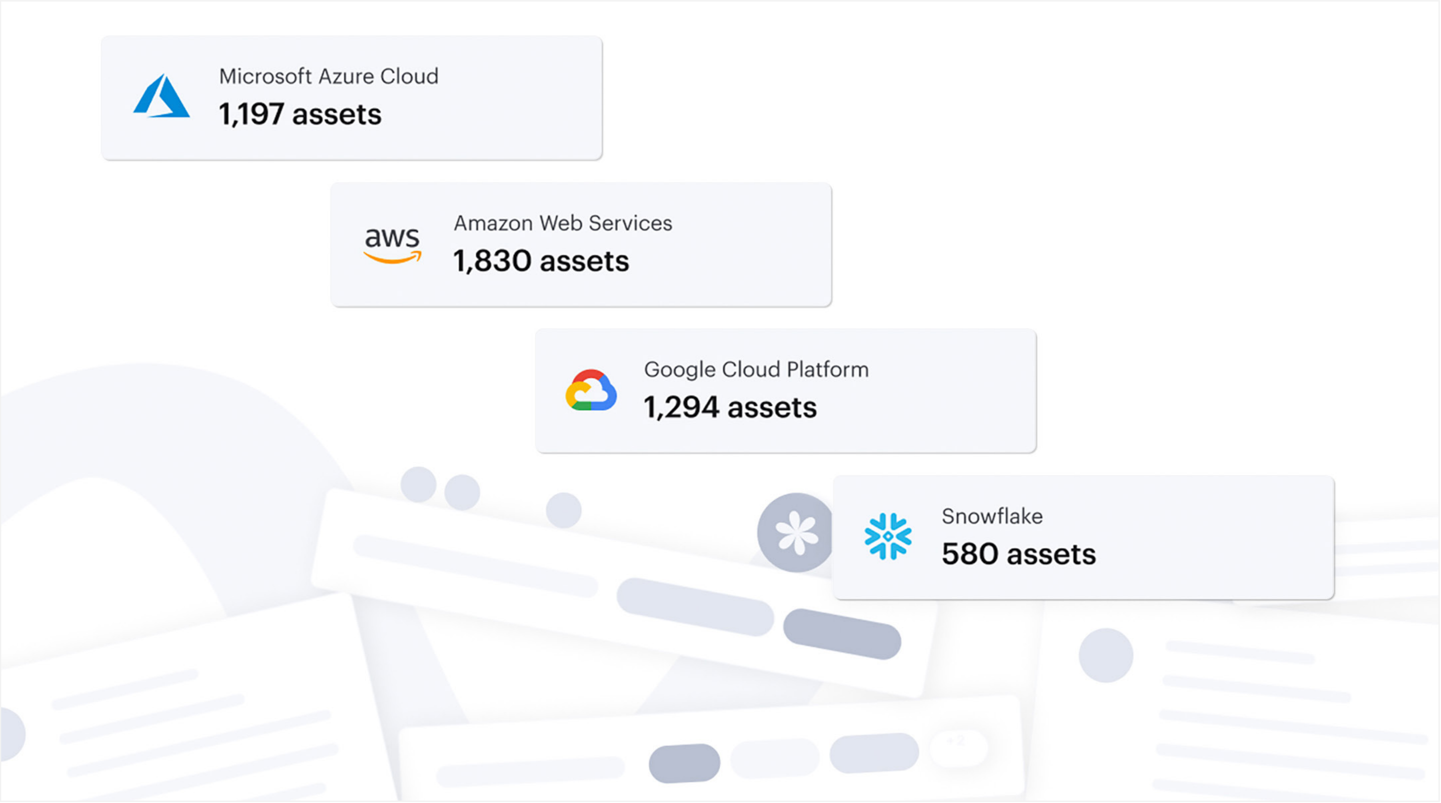

Figure 1: Data discovery

Object stores can pose significant risks due to their unstructured nature and the tendency to use them for backups, landing zones, replications, and raw data storage. Organizations might store public web assets and confidential customer information in cloud storage, increasing the likelihood of misconfigurations or human errors causing mix-ups. Virtual machines present another set of problems when, unbeknownst to security teams, they store sensitive data.

DSPM addresses these challenges by identifying all data assets in the cloud account and regularly scanning the content for sensitive records. By mapping the storage and processing of sensitive data, DSPM establishes a foundation for policy enforcement and alerts.

Classifying Sensitive Data to Prioritize Risks

Different types of sensitive data present different levels of risk and warrant specific responses. An organization might store IP addresses, PII data, credit card details, and access keys. None of these should fall into the wrong hands, but some pose a larger threat than others.

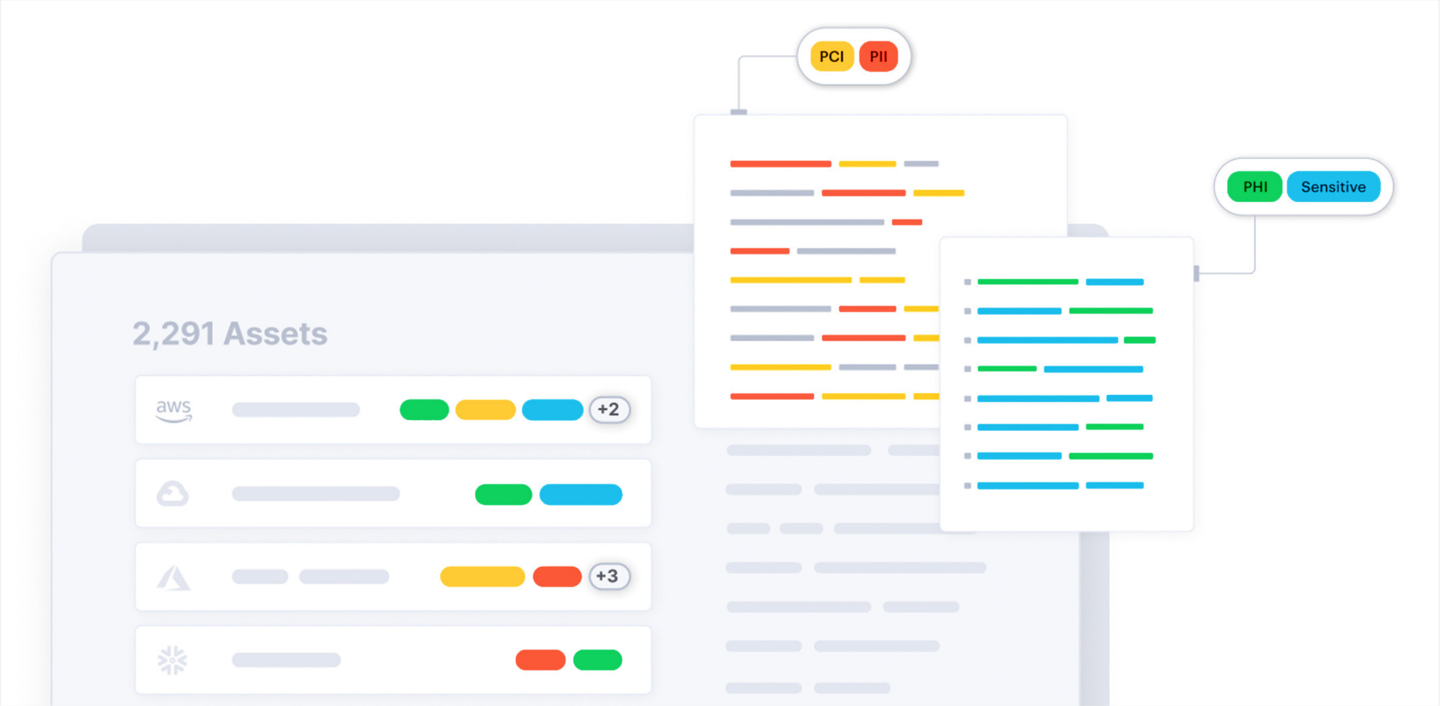

Figure 2: Classifying sensitive data

DSPM tools automatically classify each dataset in the cloud account(s), allowing security teams to prioritize policies and incident response on the most critical data assets. By prioritizing the assets containing the highest-risk data, organizations can effectively manage their data security posture and ensure that appropriate security controls for the context of the data are in place.

For example, a dataset containing PII related to named customers would likely take priority over a dataset containing aggregated, anonymized user data, making suspicious data flows involving the former high-priority issues and those involving the latter less urgent.

Static Risk Analysis Related to Sensitive Data

Once the sensitive data has been detected and classified, DSPM tools help to enforce practices meant to enhance the overall security posture related to data access—such as permissions, encrypted storage, and user management.

Monitoring and managing static risk involves examining the various security configurations and access controls associated with datastores that hold sensitive information. DSPM solutions continuously assess the cloud environment for misconfigurations, improper access controls, and other vulnerabilities that can lead to data breaches or unauthorized access. By identifying and remediating these issues, organizations can significantly reduce the likelihood of a security incident and maintain a strong data security posture.

Using DSPM capabilities, security teams can audit and adjust user permissions, identify over-privileged accounts, and enforce role-based access controls (RBAC) to limit the potential attack surface. In addition, DSPM solutions can verify that data is encrypted at rest and in transit, and that proper key management practices protect sensitive information from unauthorized access.

Understanding DSPM Through Real-World Examples

At this point, you might still be confused as to where DSPM fits in and how it differs from other solutions. The table below shows real-world risk scenarios and how a DSPM solution would help to address them.

Scenario	How DSPM Helps
Shadow backups: A database containing PII has been replicated to an unencrypted Amazon S3 bucket, which isn't managed by the central engineering organization.	Automatically discover all S3 buckets that store sensitive data, classify the sensitive data (PII, PCI, HIPAA, etc.), determine the risk level, and alert the security team.
Risky data flows: A PII record is collected through a web app, stored in CosmosDB, backed up to Azure Blob Storage, then enriched and loaded into Azure Synapse and Azure SQL for analytics and machine learning. The organization lacks visibility into the security posture of each service and the principals who have access in every step.	Map the flow of data between services and storage locations, and highlight the resources that pose a security risk- for example, due to overly permissive access rules or data duplication jobs.
Data leak from an unmanaged database: As part of an on-premises database migration, a production database is duplicated into a Windows VM. The security team is unaware that this VM is running a database, and is also unaware when a snapshot of this database is shared with a third party.	Identify that the VM is running a database and that the database contains sensitive data. When the snapshot is taken and shared, alert the SOC team in real-time so that they can take steps to prevent the exfiltration.
Snapshot exfiltration: An orphaned snapshot of an unused database, which hasn't been accessed for a long time, is now being shared with an unfamiliar account.	Identify the breach in real time and alert security teams, who can take steps to contain the attacker and prevent further data loss.
Overly broad permissions: For companies that use Google Workspace, granting permissions to Google Cloud is a matter of a few clicks. An admin gives a large group of users permissions for a specific project, then forgets to revoke it, giving dozens of principals in the organization access to PII.	Identify all the datastores that contain customer records and give security teams the means to see who has access to them. They can see that a database with sensitive information has been shared with an entire group or organization in Workspace and check whether these permissions are necessary.
Sensitive data copied by a third-party service: A data engineering team is using Fivetran to move Salesforce data into BigQuery. As part of a new technology evaluation, they use the same connection to copy a large volume of customer records from BigQuery into a Snowflake data warehouse shared with external vendors.	Map the principals, SaaS products, and vendors who can access each datastore and allow security teams to monitor sensitive data flows. The security team detects the PII being moved into Snowflake and can immediately understand the access patterns that led to the incident.

Table 1: Risk scenarios mitigated by DSPM

Permissions, Policies, and Endpoint Security Aren't Enough

With all the cybersecurity tools currently available, it makes sense to ask why you need a new category of tool. Why not rely on the existing cybersecurity stack? To answer this question, you need to understand the usage patterns of data in the modern enterprise, as well as the unique risks associated with them.

Historic approaches to securing enterprise data focused on network entry points or managing cloud configurations (CSPM) and permissions (CIEM). While implementing these remains imperative, more is needed to fully secure data in today's ecosystem.

What's Changed?

- **No endpoint to secure:** The majority of cloud data breaches never hit an endpoint. Attackers target services hosted on the public cloud, which might not be covered by enterprise VPN services. The perimeter has dissolved, and security efforts need to happen inside the cloud rather than at the entry point.
- **Impossible-to-track data usage after permissions have been granted:** With the proliferation of data, the tendency to broaden access to datasets, and the way data is replicated for analytic services, it's almost inevitable that sensitive data will end up where it doesn't belong. Permissions will struggle to cover every quick fix that becomes permanent. The same holds for cases when a developer accidentally pulls more data than needed—or forgets to delete a copy of the data sitting in a loosely-monitored S3 bucket.

DSPM does what other tools don't. By scanning data in cloud repositories where it resides, DSPM enables proactive monitoring, even of shadow data generated on the cloud or transferred between cloud services. Most importantly, it works regardless of whether the original access to the data was authorized.

The Business Logic Behind DSPM

Securing cloud data presents a technical challenge and accompanying financial risk for organizations that store sensitive data. For larger enterprises storing large amounts of customer data, the risk of compromise and the fallout of reputational and financial damages is higher.

The threat is far from theoretical. The Verizon 2023 Data Breach Investigations Report analyzed 16,312 security incidents, of which 5,199 were confirmed data breaches. "Ransomware continues its reign as one of the top Action types present in breaches, and while it did not actually grow, it did hold statistically steady at 24%."[4]

From March 2022 to March 2023, the global average cost of a data breach across industries was $4.45 million. Healthcare topped industry averages at $10.93 million.[5]

The foremost reason to improve data security is to prevent data breaches—and to reduce the amount of data exposed if one occurs.

Additional Drivers for Increasing Data Asset Scrutiny

- **Compliance:** Data security regulations affect most organizations, including data privacy laws like GDPR and CCPA, medical data legislation like HIPAA, or standards like SOC 2. Compliance with these often requires organizations to maintain an inventory of sensitive data.
- **Mergers, acquisitions, and divestitures:** During the process of buying or selling companies, enterprises need a clear picture of the data involved. Unsecure data might present a large enough risk to affect the buyer's decision. Conversely, the ability to monetize data without risking a privacy or security mishap can affect the price of the transaction.
- **Cost efficiencies:** Improving data security posture can reduce costs. This includes savings from insurance against ransomware attacks and other incidents, as well as those driven by automation processes such as policy checks, data classification, and sampling and scanning of stored data.

[4] *2023 Data Breach Investigations Report*, Verizon Business. 2023.

[5] *Global Average Cost of a Data Breach by Industry 2022*, n.d. Statista.

DSPM and the Broader Cloud Security Landscape

To fully appreciate the vital role of DSMP requires an understanding of where it sits within the cybersecurity ecosystem.

DSPM vs. CSPM

How does DSPM differ from cloud security posture management (CSPM), which the industry, until recently, considered the primary approach for protecting cloud data assets?

CSPM solutions focus on protecting the infrastructure. Policies center on reviewing data replication rules, fine-tuning access control, and finding weaknesses in cloud infrastructure and design. While these measures are vital, scanning the data—something CSPM is not built for–is also vital.

DSPM looks beyond the policy level to the content of the data. By scanning and classifying enterprise data, it allows organizations to see where sensitive data exists and how it's being utilized. It also helps prioritize the long list of discovered issues, preventing alert fatigue, which can cause issues to be ignored.

DSPM vs. Other Cloud Security Tools

We've talked about the differences between DSPM and CSPM. But how does DSPM compare to other types of cloud data security solutions?

Data loss prevention (DLP) tools were designed for data exfiltration from the endpoint and not from the cloud. DSPM tools offering real-time DDR capabilities are a form of cloud DLP that complement traditional DLP solutions.

Cloud access security brokers (CASBs) help enforce security policies between data consumers and SaaS applications. They don't cover data after it's stored on IaaS, PaaS, or DBaaS, which is where DSPM comes in.

Native solutions offered by public cloud vendors (AWS, Azure, Google) don't support multicloud environments and are often limited in coverage and functionality (e.g., covering only one type of service or database). DSPM provides holistic coverage, even in multicloud environments.

Advantages of DSPM

- **Better visibility into where sensitive data lives:** DSPM solutions scan cloud data repositories, discover sensitive data, and classify it. Creating an accurate map and inventory of the organization's data assets helps organizations to understand where sensitive data is stored, who is accessing the data, and where it's going.
- **Identify data risks:** Static risk analysis identifies data lacking protection and prevents the misuse of data assets. The types of checks performed by DSPM include ensuring that data is encrypted and logging is enabled for any situation where sensitive data can be accessed.
- **Policy controls:** DSPM solutions provide a policy engine supported by a deep data threat model. They detect real-time risks, allowing for immediate remediation to stop a potential breach.

Data Security Coverage by Solution

Solution	Focus	How DSPM Complements or Differs
CSPM	Cloud infrastructure protection and configuration management	DSPM scans and classifies actual data, providing visibility into where sensitive data is located and how it's being used. It helps prioritize discovered issues and prevents alert fatigue.
DLP	Data exfiltration prevention from the endpoint	DSPM tools with real-time data detection and response (DDR) capabilities can be seen as a form of cloud DLP, focusing on data protection in cloud environments.
CASB	Enforcing security policies between data consumers and SaaS applications	DSPM covers data stored in IaaS, PaaS, and DBaaS, providing comprehensive data protection across cloud services.
Native Cloud Solutions	Data protection and security within a specific cloud provider's ecosystem	DSPM provides holistic coverage, including in multicloud environments, and offers more extensive coverage and functionality.

Table 2: At-a-glance comparison of security solutions

Why DSPM Is Dominating the Conversation?

The public cloud has transformed the way we work with data. Organizations no longer rely on monolithic databases or DevOps platforms. Instead, developers leverage the cloud's elasticity to adopt a range of tools, as well as microservices. These new paradigms give product and analytics teams more space to innovate and iterate quickly. At the same time, they create risks when it comes to sensitive data. DSPM vendors have emerged to address the evolving needs of enterprise data security.

Let's examine the factors moving data-centric security to the top of CISO priorities.

The Breakdown of the Enterprise Data Warehouse

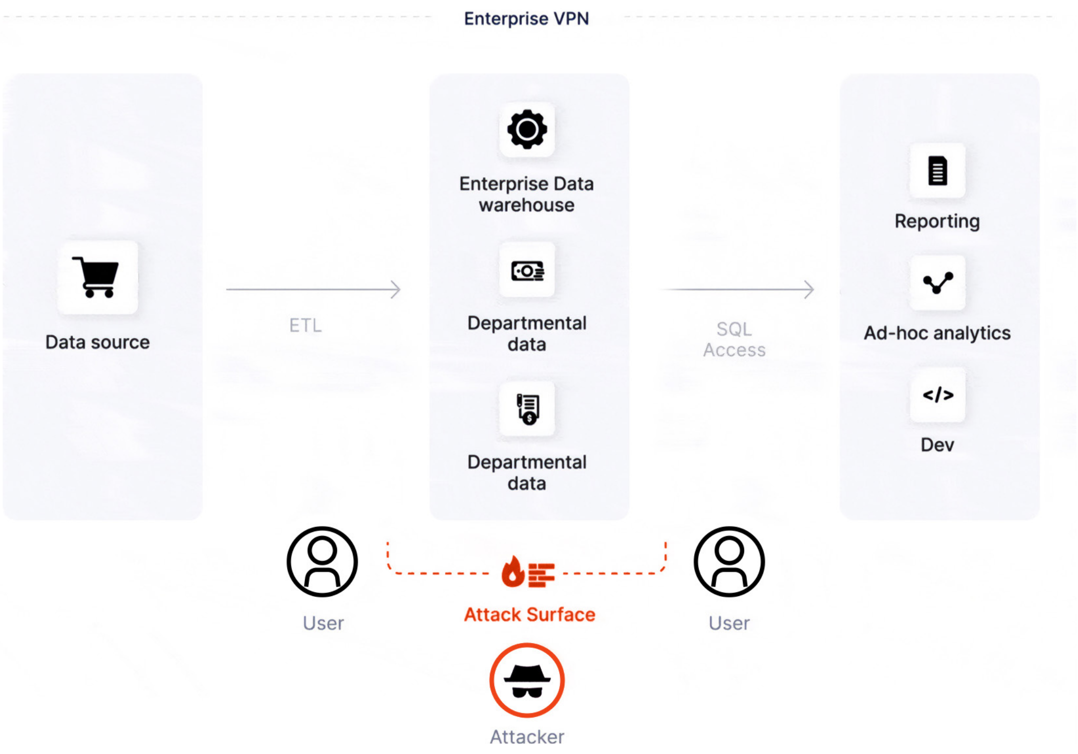

Figure 3: The attack surface of the legacy architecture includes the enterprise VPN, the enterprise data warehouse, and the departmental data.

From the late 1980s through the early 2000s, data largely resided in a single enterprise-wide data warehouse (EDW), such as Oracle or Teradata systems. Security teams had a well-defined attack surface—securing the data warehouse. Access to data was often through DBA teams, which could maintain strict oversight.

The contemporary approach, in contrast, emphasizes data democratization, broader access, and the use of various best-in-breed tools to tackle specific data challenges. While this fosters data-driven decision-making, it also expands the attack surface across dozens, if not hundreds, of datastores, increasing security complexities.

It's uncommon today for large organizations to rely solely on a single EDW. The flexibility of the cloud makes it easy to spin up new services and retain larger amounts of raw data. Modern cloud architectures often use lower-cost object storage, such as Amazon S3, for storing raw data, which can then be processed in an array of databases or analytic services for various use cases in the organization.

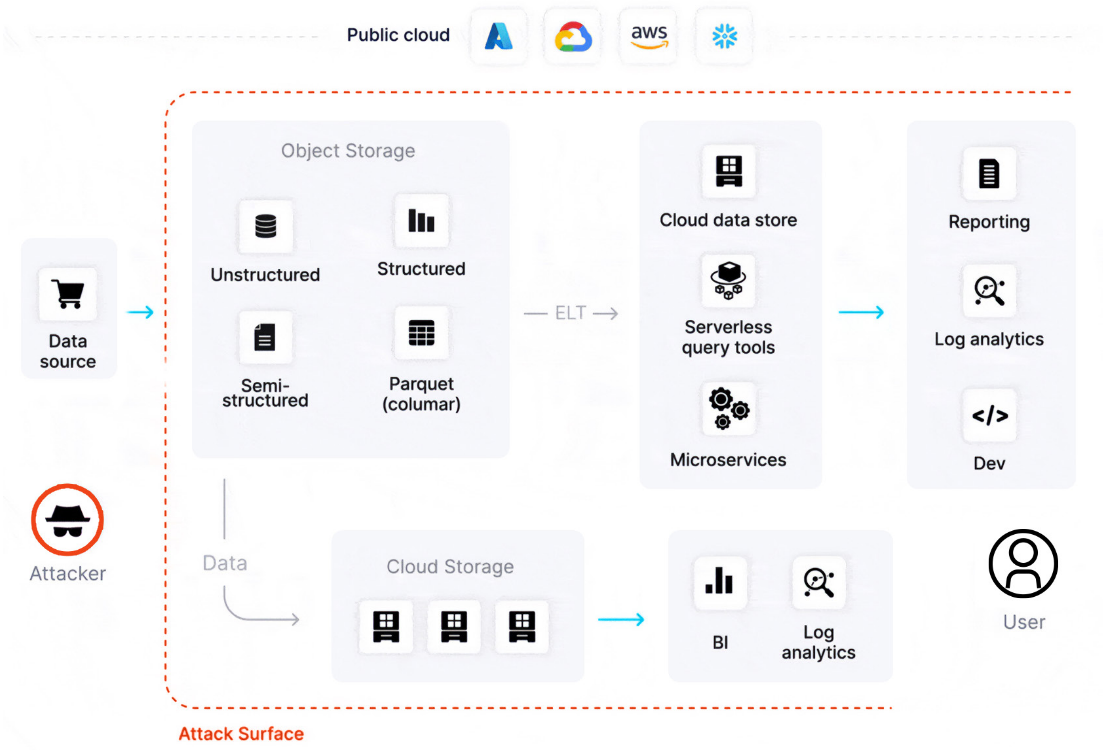

Figure 4: While multicloud architectures generally have a larger attack surface than legacy architectures due to the increased complexity of dispersed entry points, successful attacks require breaching multiple systems.

Let's look at a scenario in a financial services organization where multiple data storage and processing systems are utilized. The company stores its raw transaction log in object storage while copying a subset of the data to Snowflake for analytics purposes. They also move some logs into Elasticsearch for application troubleshooting and provide data science teams with access to the raw data to run Spark ETL and machine learning jobs. For each use case, data is copied and moved, adding potential locations where sensitive data might end up.

Microservices-Based Development

It's not just storage that's distributed. Modern software engineering favors breaking monolithic applications into microservices—smaller applications or pieces of code that communicate via APIs.

Microservices give developers flexibility while freeing them from an overreliance on DevOps. Benefits aside, though, each microservice has data assets assigned to it, leading to a proliferation of data copies with minimal oversight.

A great deal of modern engineering work involves handling and analyzing data. It's almost inevitable that developers will move or replicate sensitive data in the process of writing new code.

Multicloud Architectures

Exacerbated by the adoption of multicloud environments, the ease of moving data between services leads organizations to adopt tools from different cloud providers. The same dataset, for example, might find itself in Amazon Aurora and Azure Synapse due to different teams needing to run different SQL queries or optimize costs. The same dataset, in other words—to be clear—is managed by different tools.

As data moves between clouds, tracking lineage and classification takes on greater complexity. Native tools offered by the public cloud providers are limited to that specific cloud. The likelihood for sensitive data to seep into unmonitored corners rises. In such circumstances, with inherent gaps in visibility, establishing effective oversight can be extremely challenging.

> “DDR solutions use real-time log analytics to monitor cloud environments that store data and detect data risks as soon as they occur.

Improving DSPM Solutions with Data Detection and Response (DDR)

The DSPM capabilities we've discussed to this point refer primarily to static risk—finding sensitive data, classifying it, and reviewing the access controls and configurations applied to it.

To maintain an effective data security posture, though, you need to continually monitor and analyze data access patterns and user behavior. Data detection and response (DDR) does just that.

DDR provides real-time monitoring and alerting capabilities to help security teams quickly detect and respond to potential threats and suspicious activities—while it prioritizes issues that put sensitive data at risk. By leveraging machine learning algorithms and advanced log analytics, DDR can identify anomalies in user behavior and access patterns that potentially indicate a compromised account or insider threat.

A Closer Look at Data Detection and Response (DDR)

DDR describes a set of technology-enabled solutions used to secure cloud data from exfiltration. It provides dynamic monitoring on top of the static defense layers provided by CSPM and DSPM tools.

With today's organizations storing data across various cloud environments—PaaS (e.g., Amazon RDS), IaaS (virtual machines running datastores), and DBaaS (e.g., Snowflake)—it isn't feasible to monitor every data action. DDR solutions use real-time log analytics to monitor cloud environments that store data and detect data risks as soon as they occur.

How DDR Solutions Work

DDR solutions incorporate DSPM capabilities to discover and classify data assets, identify risks such as unencrypted sensitive data or data sovereignty violations, and prioritize remediation by data owners or IT. Once sensitive data assets are mapped, DDR solutions monitor activity through cloud-native logging available in public clouds, generating event logs for every query or read request.

The DDR tool analyzes logs in near real-time, applying a threat model to detect suspicious activity, such as data flowing to external accounts. Upon identifying new risks, DDR issues alerts and suggests immediate actions. These alerts are often integrated into SOC or SOAR (security orchestration, automation and response) solutions for faster resolution and seamless alignment with existing operations.

DDR Use Cases

To envision the types of incidents a DDR solution addresses, consider a few examples seen among our users.

- **Data sovereignty issues:** Legislation from recent years creates obligations to store data in specific geographical areas (such as the EU or California). DDR helps detect when data flows to an unauthorized physical location, preventing compliance issues down the line.
- **Assets moved to unencrypted/unsecure storage:** As data flows between databases and cloud storage, it can make its way to an insecure datastore (often the result of a temporary but forgotten workaround). DDR alerts security teams to this type of movement.
- **Snapshots and shadow backups:** Teams face increasing pressure to do more with data, leading to the prevalence of shadow analytics outside approved workflows. DDR helps find copies of data stored or shared in ways that may cause breaches.

How Does DDR Fit into the Cloud Data Security Landscape?

DDR vs. CSPM and DSPM

- Cloud security posture management (CSPM) is about protecting the posture of the cloud infrastructure (such as overly generous permissioning or misconfiguration). It doesn't directly address data—its context and how it flows across different cloud services.
- Data security posture management (DSPM) protects data from the inside out. By scanning and analyzing stored data, DSPM tools identify sensitive information such as PII or access codes, classify the data, and evaluate its associated risk. This process provides security teams with a clearer picture of data risk and data flow, enabling them to prioritize cloud assets where a breach could cause the most damage.

While DSPM offers more granular cloud data protection, both CSPM and DSPM are static and focused on posture. They allow organizations to understand where risk lies but offer little in terms of real-time incident response.

In contrast, DDR is dynamic. It focuses on data events happening in real time, sending alerts, and giving security teams a chance to intervene and prevent significant damage. DDR monitors the specific event level, whereas other solutions look at configurations and data at rest.

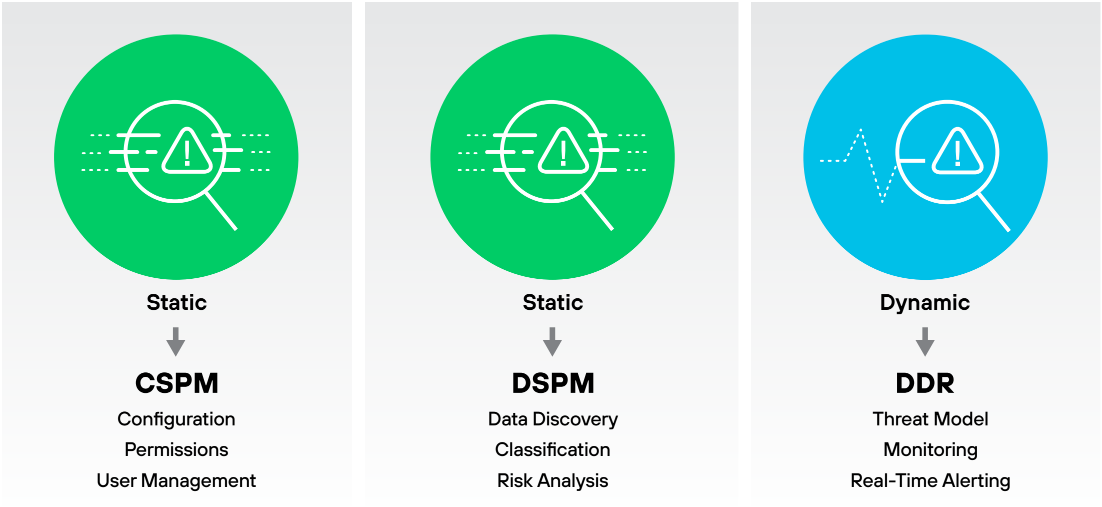

Figure 5: DDR vs. CSPM and DSPM

A Potential Situation

Consider a scenario where an employee has authorized, role-based access to a database containing customer data. The employee plans to leave the company and, before notifying their managers of their intention to leave, copies the database to their personal laptop to take to the next company.

In this example, permissions allow the employee to access the database—and yet, a major exfiltration event is unfolding. A DDR solution with a well-calibrated threat model detects the unusual batch of data (and other irregularities) contained in this export. The DDR tool sends an alert to the security team and provides full forensics—pinpointing the exact asset and actor involved in the exfiltration. Saving critical time, the security team intervenes before any real damage is done.

Does the CISO Agenda Need an Additional Cybersecurity Tool?

DDR provides mission-critical functionality missing from the existing cloud security stack. When agents aren't feasible, you need to monitor every activity that concerns your data. DDR protects your data from exfiltration or misuse, as well as from compliance violations. By integrating with SIEM and SOAR solutions, enabling teams to consume alerts in one place, DDR helps reduce operational overhead.

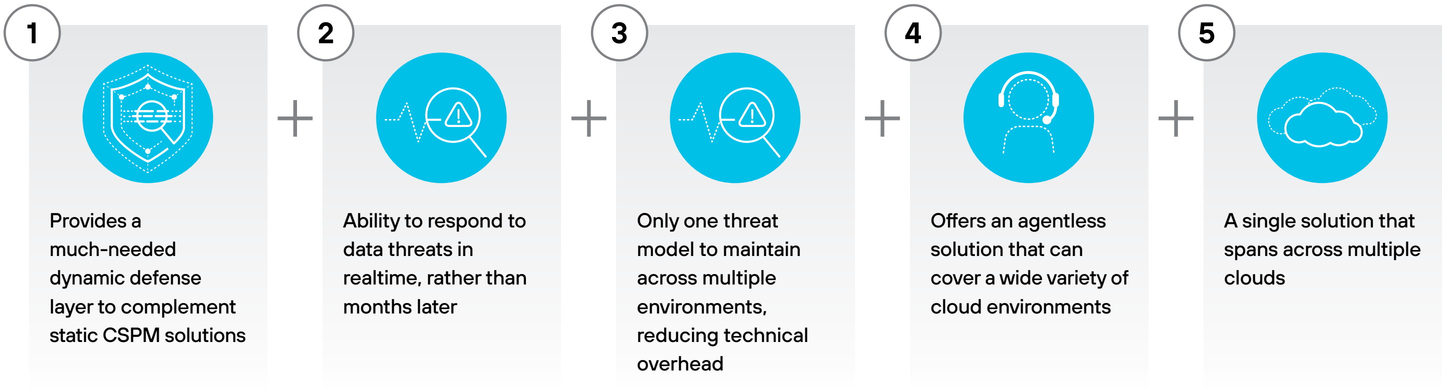
1
Provides a much-needed dynamic defense layer to complement static CSPM solutions
+
2
Ability to respond to data threats in realtime, rather than months later
+
3
Only one threat model to maintain across multiple environments, reducing technical overhead
+
4
Offers an agentless solution that can cover a wide variety of cloud environments
+
5
A single solution that spans across multiple clouds

Figure 6: Five key benefits of data detection and response

The Need for Agentless DLP

Monitoring data assets in real time might seem obvious, but organizations, for the most part, lack an adequate way to protect sensitive data. In the traditional, on-premises world, work was mainly done on personal computers connected via an intranet to a server. Security teams monitored traffic and activity by installing agents (software components such as antivirus tools) on every device and endpoint that had access to organizational data.

But you can't install an agent on a database hosted by Amazon or Google or place a proxy in front of thousands of datastores. The move to cloud infrastructure requires new approaches to data loss prevention (DLP).

The industry gravitated toward static solutions geared toward improving the security posture of cloud datastores (CSPM, DSPM) by detecting misconfigurations and exposed data assets. But the challenge with data flow hadn't been addressed until DDR.

When Static Defense Layers Aren't Enough: Lessons from a Breach

The 2018 Imperva breach began with an attacker gaining access to a snapshot of an Amazon RDS database containing sensitive data.[6] The attacker used an AWS API key stolen from a publicly accessible, misconfigured compute instance.

Would CSPM and DSPM have prevented the breach?

A CSPM solution could identify the misconfiguration, and DSPM could detect sensitive data stored on the misconfigured instance. Neither tool, though, would have been able to identify the unusual behavior once the attacker had gained access that appeared legitimate.

And as it unfolded in 2018, the Imperva breach wasn't discovered for 10 months, via a third party. The attacker had exported the database snapshot to an unknown device—and, all the while, the unaware company couldn't notify its users that their sensitive data had been leaked.

A DDR solution would have addressed the gap by monitoring the AWS account at the event log level. Potentially identifying the attack in real time, it would have alerted internal security teams, allowing them to respond immediately.

Supporting Innovation Without Sacrificing Security

The cloud is here to stay, as are microservices and containers. As cybersecurity professionals, we can't prevent the organization from adopting technologies that accelerate innovation and give developers more flexibility. But we need to do everything we can to prevent data breach.

DSPM with DDR offers critical capabilities previously missing in the cloud security landscape—data discovery, classification, static risk management, and continuous and dynamic monitoring of complex, multicloud environments. Providing organizations with the visibility and control necessary to effectively manage their data security posture enables organizations to catch incidents earlier, averting or minimizing disastrous data loss.

[6] *Imperva: Data Breach Caused by Cloud Misconfiguration*, n.d. Threatpost.com.

Prisma Cloud by Palo Alto Networks

Prisma® Cloud secures your applications from Code to Cloud™ across multicloud environments. The platform delivers continuous visibility and threat prevention throughout the application lifecycle, including zero day threats. With Code to Cloud coverage that encompasses code, infrastructure, workloads, data, networks, web applications, and APIs security, Prisma Cloud is the only platform that addresses your security needs at every step in your cloud journey. With over 4 billion cloud assets secured, you can trust Prisma Cloud to protect your cloud at any scale. Prisma Cloud enables security and DevOps teams to effectively collaborate to accelerate secure cloud native application development and deployment.

For more information, check out paloaltonetworks.com/prisma/cloud

3000 Tannery Way
Santa Clara, CA 95054

Main:	+1.408.753.4000
Sales:	+1.866.320.4788
Support:	+1.866.898.9087

www.paloaltonetworks.com