rubrik

EBOOK

A Buyer's Guide to Data Security Posture Management (DSPM) Solutions

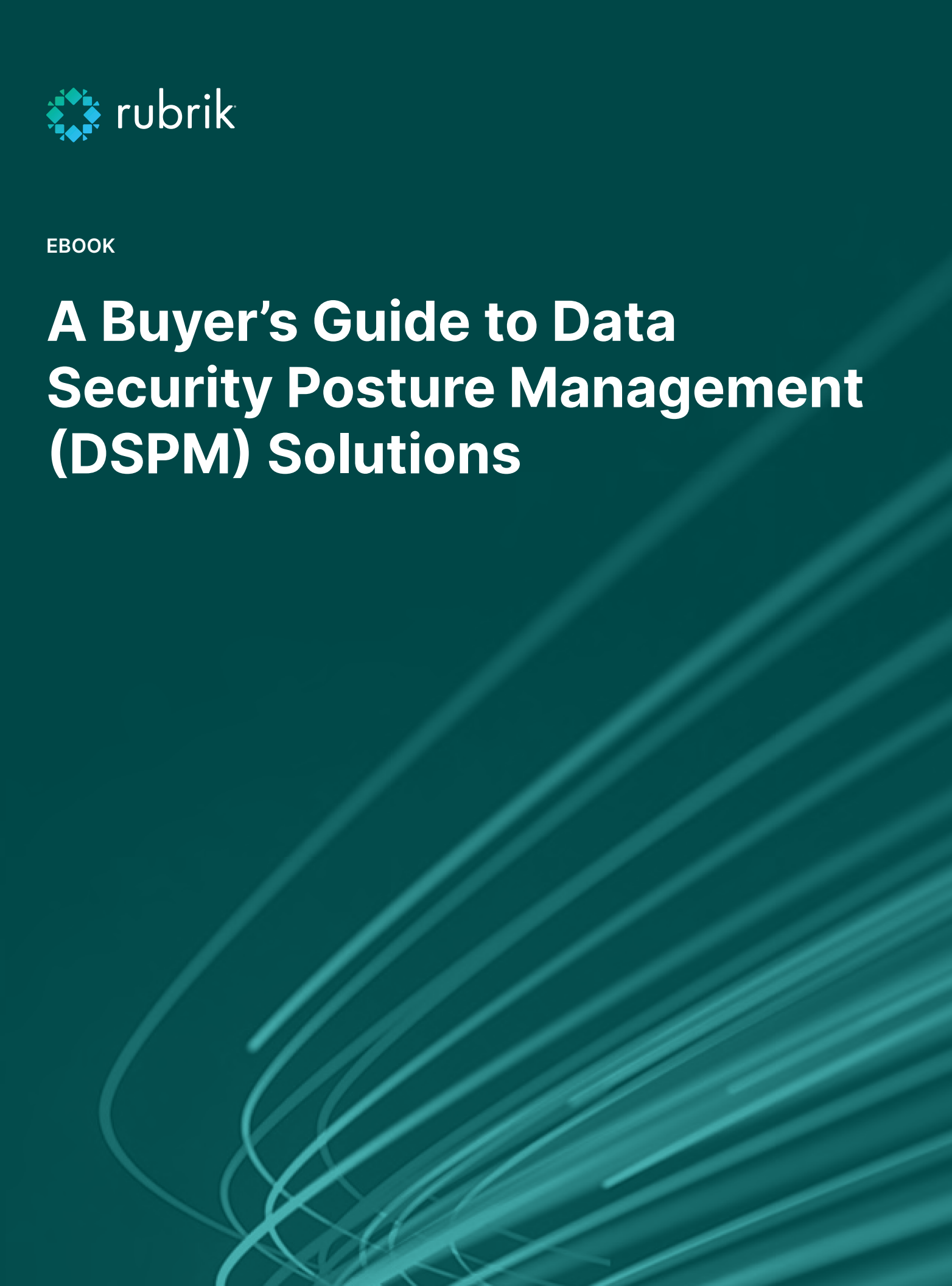


Table of Contents

This guide discusses the changes in the world that have precipitated the need for data security posture management (DSPM), what it is, how it compares to all other approaches to data security in the cloud, and what to look for when evaluating a solution.

CHAPTER 1 | DATA SECURITY CHALLENGES IN THE CLOUD

With the widespread adoption of digital transformation and cloud migration, businesses are generating an unprecedented amount of data. The cloud's scalability and flexibility enable companies to share data easily, which has become a driving force for business innovation and organizational agility. However, data democratization, or the process of making data accessible to a wider range of employees, has also presented new challenges for security teams, including data privacy and governance concerns.

Unfortunately, the same activities that data democratization has enabled, those that fully utilize data and create the biggest competitive advantages for businesses, are also the ones that pose the biggest risks. As agile application development and innovation activities increase, data proliferation naturally follows. And unfortunately when data proliferates, security controls don't travel with it.

As this democratization process happens, it is very common for sensitive, proprietary and highly regulated data to get copied, modified, shared and moved across multiple cloud data stores without any oversight by data security teams. In fact, it only takes seconds for an application developer or data scientist, to:

- Accidentally making a file publicly accessible in a cloud storage location like S3
- Copy a database to a test environment without proper protection and forget to delete it afterwards
- Create a snapshot of an RDS, move on to other projects, and forget to delete it
- Delete sensitive data from a managed datastore without realizing that versioning is enabled, allowing deleted data to still be accessed through a restore process

These practices can occur repeatedly throughout an entire organization, undetected by even the most diligent security teams. This new reality results in the proliferation of "shadow data" and the emergence of a new threat vector called the "innovation attack surface," making critical security, privacy, and governance functions in the cloud increasingly complex and challenging to perform.

WHAT IS THE INNOVATION ATTACK SURFACE?

Traditionally, we viewed a business's "attack surface" from the outside looking in: anything left unprotected that was exploitable by external forces. In an on-premises world, it only took protecting each static asset—usually a physical piece of hardware—to mitigate risk in this traditional environment.

But the innovation attack surface looks entirely different. It is a new threat vector that most organizations unconsciously accept as the cost of doing business. In contrast to traditional attack surfaces determined by external forces (including bad internal actors) seeking to exploit vulnerabilities to gain illicit access to sensitive information the innovation attack surface results from the massive, non-contiguous patchwork of accidental risk created by the organization's most highly active cloud users - its developers and data scientists, when leveraging data for innovation.

Malicious actors are quickly becoming aware of this novel attack surface and are actively seeking to exploit it. Therefore it is imperative that security teams remain vigilant in safeguarding an organization's most critical data to prevent unauthorized access and prevent potential data breaches.

WHAT IS SHADOW DATA?

As data democratization becomes more widespread, a new problem has emerged: shadow data. This refers to data that has been unknowingly created, copied, backed up, or stored in a data store without the knowledge of security or IT teams, leaving it overexposed, unprotected, and unmonitored.

The problem has become significant for organizations, as neither security nor IT teams know what data they have or where it is stored. Therefore, they are unable to secure or govern it. In fact, a recent report, "The State of Public Cloud Security 2023," revealed that 93% of respondents were either fully or extremely concerned about shadow data.

For organizations operating in the cloud, the innovation attack surface and shadow data present a significant challenge and play a big part in the significant rise in data breaches and exfiltrations. Businesses must recognize these issues, as a single breach costs an average of $9.44 million in the United States, and that price tag does not include their public image, trust, and customer loss.

SECURING DATA IN THE CLOUD DEMANDS A NEW APPROACH

Despite companies' continued investment in security technologies, the number of data breaches continues at an alarming rate. In 2023 alone, there were 3,205 publicly reported data compromises that impacted an estimated 3.5 million people (Identity Theft Resource Center). This alarming trend can be attributed to the fact that legacy data security solutions are not designed to adapt to modern, more dynamic environments, as explained in Chapter 2. As a result, organizations need to implement more agile and adaptable security solutions that can keep pace with the rapidly evolving threat landscape.

The shortcomings of existing data security approaches, coupled with the shortage of skilled cloud technology professionals, have led to a "security execution gap." This gap is best described as the growing divide between the activities that contribute to business innovation and the security activities and expertise necessary to safeguard the business.

For organizations looking to empower their developers and data scientists to innovate more quickly and securely than their competitors, bridging the security execution gap must be a top priority. This can only be achieved by investing in data security solutions that are agile enough to keep pace with the speed of innovation.

This guide will explore the most effective approach to closing this gap and securing data in the cloud: data security posture management (DSPM).

CHAPTER 2 | WHAT IS DSPM?

What is Data Security Posture Management (DSPM)? Fundamentally, DSPM refers to the processes, policies, and technologies used to protect sensitive data and ensure compliance in cloud environments at scale and with automation. It is rapidly evolving as a security solution category in response to the growing threat of the innovation attack surface. In fact, DSPM was named an "on the rise" technology by Gartner in their 2022 Hype Cycle for Data Security.

> Gartner's definition:
>
> According to a Gartner® report, data security posture management (DSPM) provides "visibility as to where sensitive data is, who has access to that data, how it has been used, and what the security posture of the data store or application is."[1]

The most challenging aspect of data security is locating and securing cloud data. This is where cloud-native data security posture management (DSPM) comes in by delivering autonomous and continuous data discovery, classification, and protection across multiple cloud platforms. While other solutions claim to offer similar levels of security, they often fail to deliver. We will explore why.

DSPM VS. OTHER DATA SECURITY APPROACHES

When it comes to DSPM alternatives, there are many disadvantages, especially when it comes to determining which data assets should be scanned. It is always the organization's responsibility to locate these assets manually and then configure connectors for them. This process is not only time-consuming, labor-intensive, and complex, but it is also completely ineffective in the cloud, leaving large amounts of sensitive data unaccounted for.

Other alternatives are often only compatible with certain cloud platforms and support limited types of data assets. Additionally, many DSPM alternatives remove data from the organization's environment in order to scan it, ultimately increasing the risk.

While some organizations start with manual or homegrown methods or rely on CSP-native or legacy security tools, they soon realize these approaches are not sufficient, sustainable, or scalable enough to secure all their cloud data.

1 Gartner, Hype Cycle™ for Data Security, 2022, Brian Lowans, 4 August 2022. GARTNER and Hype Cycle are registered trademarks and service marks of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.

A COMPARISON OF OTHER APPROACHES VS. CLOUD-NATIVE DSPM

Manual Approaches

- Tedious and resource-intensive
- Inadequate for dynamic, rapidly changing environments (quickly becomes obsolete<
- Prone to human error
- Manual identification and enforcement of policy violations
- Unable to detect shadow data

Cloud-native DSPM

- Leverages cloud-native APIs to discover, scan, and classify sensitive and shadow data
- Enables continuous monitoring of the environment without requiring human intervention
- Continuously evaluates the security posture of each data asset
- Automatically identifies and alerts on policy violations

Homegrown Tools

- Diverts engineering resources and time from strategic security initiatives
- Requires advanced knowledge and ongoing proficiency in complex cloud environments
- Struggles to keep up with the rapid pace of cloud innovation
- Often fails to detect shadow data

Cloud-native DSPM

- Tailored for securing data in cloud environments
- Optimized for the dynamic and fast-paced nature of the cloud
- Ensures cloud data is protected, freeing up teams to innovate
- Continuously adapts to evolving data security threats
- Autonomously discovers both known and shadow data

CSP-Native Tools
(AWS Macie, Azure Purview, GCP Cloud DLP)

- Coverage is limited to designated cloud environment
- Scanning is limited to configured assets, missing shadow data
- Limited support for certain data asset types
- Point-in-time scanning results, no continuous monitoring
- No data security posture monitoring or actionable remediation guidance
- Expensive “per use” licensing model

Cloud-native DSPM

- Consolidates security and contextual metadata from multi-clouds into one console
- Autonomously discovers and classifies all cloud data, across all cloud platforms
- Provides technical recommendations for fixing security and compliance violations
- Evolves with cloud innovation and data security
- Continuously monitors for new assets and data changes

Legacy Data Security Solutions

- Built for on-premises environments only
- Manual connection required for each data asset, including access credentials
- Connector-based, unable to scale in the cloud or discover shadow data
- Additional security risks with data removal from environments
- High operational costs due to ongoing manual connections
- Leaves significant data pockets uncovered

Cloud-native DSPM

- Cloud-native and designed for the dynamic nature of the cloud
- Continuously monitors the cloud ecosystem for new or modified data assets
- Autonomously discovers and classifies all cloud data
- Secures cloud data without removing it from the environment
- API-only architecture provides asynchronous, zero-performance impact
- Rapid deployment and time-to-value

Cloud Infrastructure Security Tools (CSPM, CNAPP, etc.)

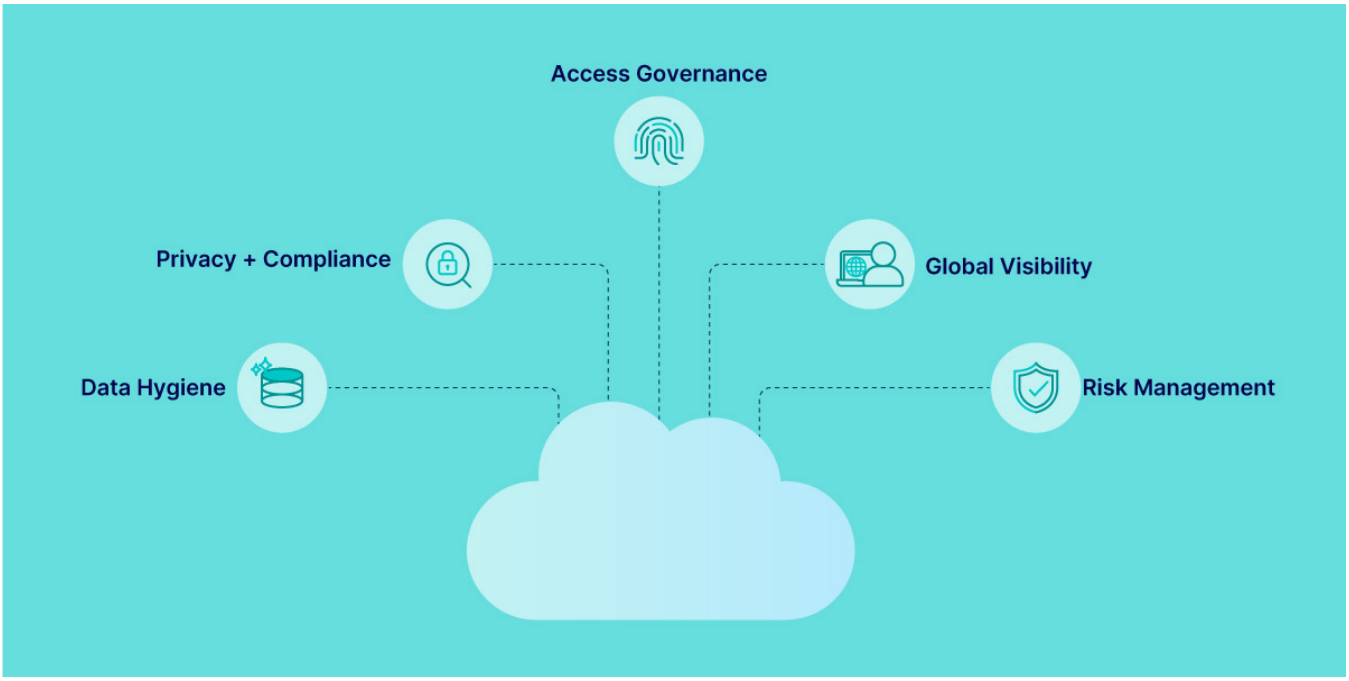
- Infrastructure security team is the primary user, not data security
- Provides basic data discovery with limited context and no focus on privacy, compliance, or governance requirements
- Does not discover or alert on misplaced or redundant data
- Increases risk by moving data outside your cloud environment

Cloud-native DSPM

- Designed for data security teams that prioritize security, governance, and privacy requirements independent of infrastructure
- Provides granular data insights, including location, ownership, access control, posture status, and usage
- Supports full data security requirements, including governance and privacy policies, identifying over-exposed, under-protected, misplaced, and redundant data
- Best-in-class DSPM ensures data remains within your cloud environment

CHAPTER 3 | WHAT TO LOOK FOR IN A DSPM

Having gained an understanding of why DSPM is crucial and the shortcomings of other data protection solutions in the cloud, the next step is to evaluate various DSPM offerings. It is critical to select the right solution, as an effective DSPM solution should help your organization navigate the risks associated with cloud data, facilitate innovation, and reduce the attack surface, all while enabling data-driven growth. To accomplish this, your DSPM solution should include the following capabilities:



Global data visibility – gain a unified view of all sensitive data assets, including each one's location, ownership, access control, posture status, and usage.

Data hygiene – Address and remediate misplaced, redundant, and obsolete data. This includes identifying and preventing data exposure and unnecessary duplication, purging outdated or irrelevant data, and ensuring that policies are in place to monitor data hygiene continuously.

Data security risk management – Identify and address security risks associated with overexposed, unprotected, and misplaced data through detection, prioritization, and remediation.

Data access governance – Identify all internal and external users, roles, and resources with access to sensitive data. Then, monitor and control each user's access to sensitive data based on their roles and responsibilities. This process ensures that only authorized users have access to sensitive assets.

Privacy & compliance – Detect and remediate violations of data privacy regulations and industry standards (GDPR, PCI DSS, etc.), then generate audit-ready compliance reports.

To understand how DSPM solutions accomplish these, we can break down their primary functions into three criteria categories—each enabled by several key features. You should look for a solution that can address as many of these features as possible.

CRITERIA 1 – DATA DISCOVERY AND CLASSIFICATION CAPABILITIES

The first step to ensuring data security in the cloud is to understand which data assets you own, where they reside, who owns them, and who can access them. A holistic data discovery and classification process is essential for improving security, and implementing effective data governance, which ultimately reduces risk and simplifies compliance.

However, the cloud's fast-moving and ever-changing nature makes this process highly complex. Security teams must monitor countless moving parts simultaneously to discover and classify everything within their organization's purview. To overcome these challenges, the most effective DSPM solutions should offer the following capabilities:

AUTONOMOUS AND CLOUD-NATIVE DATA DISCOVERY & CLASSIFICATION

Many data security solutions require users to manually connect to each data asset, including the need for access credentials. In a cloud environment, this is virtually impossible due to the sheer volume of data stores and pace of change. Instead, a DSPM should be cloud-native and utilize the CSP's APIs to autonomously scan your environment without requiring intervention from your team. It should automatically detect and classify all data, including the elusive "shadow data," stored within your cloud environment without any prior knowledge, without bothering anybody or asking for connection information.

CONTINUOUS DATA DISCOVERY & CLASSIFICATION

A developer or data scientist can create or duplicate entire data stores in just a few clicks. A DSPM solution cannot rely solely on point-in-time snapshots to capture these frequent changes. Instead, the platform should continually monitor your environment for modifications and automatically scan for new data assets, changes to existing accounts, as well as identifying entirely new accounts.

BREADTH OF COVERAGE

A DSPM must function across multi-cloud environments, extending to all major cloud service providers. It should also be able to read from various databases, data pipelines, object storage, disk storage, managed file storage, data warehouses, lakes, and analytics pipelines (S3 buckets, BigQuery, Redshift, Storage Container, EBS, RDS, DynamoDB, etc.) both managed and self-hosted. Make sure your DSPM provider goes wide and beyond just basic object storage.

DEPTH OF COVERAGE

The DSPM solution needs to be able to classify sensitive data in structured, semi-structured, and unstructured formats. Depth of coverage means that the DSPM has support for different variations of technologies in each subset of data assets, including different types of self-hosted databases with unique configurations (such as an EC2 running an MSSQL engine). It should work alongside several self-hosted databases with different configurations and file formats. In addition, it must support different kinds and formats of files like JSON, Office, Avro, Parquet, and others.

INTELLIGENT CLASSIFICATION

The DSPM should also be able to classify data accurately with depth and breadth. It should automatically identify sensitive data with minimal input and return contextual results with low false-positive and false-negative rates. Multi-step, intelligent classification goes beyond regular expression matching (reg-ex) by evaluating the context of the data type to ensure accuracy and lower false positives.

LOW COST OF SCANNING

Scanning data in a cloud environment is a computationally intensive process, and any errors can significantly increase cloud costs. To optimize the scanning process, an effective DSPM solution should consider context and adjust its scanning strategy accordingly. This means it should prioritize efficiency by skipping redundant data and using sampling techniques when appropriate. Additionally, it should optimize costs by adopting the mindset that if data hasn't changed, there is no need to rescan it.

SECURE BY DESIGN

When selecting a DSPM solution, it's essential to consider whether the solution keeps your data safe while scanning. Some security tools copy data from your environment into the vendor's, thereby increasing your attack surface. Thus, choosing a DSPM solution that scans data within your environment is crucial. Additionally, look for a solution that comes with advanced cloud features, such as serverless functions that leverage APIs, to enhance your security posture.

In a cloud environment, data discovery and classification should be considered an ongoing process. The dynamic nature of data in cloud environments means that continuous global data visibility is necessary.

This approach monitors constantly changing data structures and storage locations, ensuring that the data security team has a unified view of all sensitive, proprietary and regulated data assets, including each one's location, ownership, access control, posture status, and usage.

CRITERIA 2 – DATA SECURITY POLICIES FRAMEWORK

If you want to automate your data security policy verification process, you need to start with a comprehensive data-centric policy framework. This framework ensures that you're addressing all categories of sensitive data violations and promotes good data hygiene practices, making it easier to maintain compliance and protect your organization's sensitive, proprietary and regulated data.

To enable an effective policy framework, select a DSPM solution that offers:

BOTH BUILT-IN AND CUSTOM POLICIES

Your DSPM solution should offer built-in and custom policies that address all categories of sensitive data violations, such as overexposed, unprotected, misplaced, redundant, and non-compliant data. These policies must support a wide array of data asset types, environments, and security posture controls across your cloud environment. Adhering to these data-centric policies not only promotes strong data security but also ensures data hygiene by cleaning up and preventing duplicate and unnecessary data creation.

The best DSPMs enforce both organizational policies and widely accepted industry and regulatory standards. For organizations just starting out, having a large number of out-of-the-box policies matched to various frameworks is extremely helpful. For companies with existing, well-documented data security policies, the ability to configure these policies into code that can be automatically verified is essential. The framework should also automatically notify designated individuals of any policy violations. Clear explanations and the ability to drill down to see evidence of the violation facilitate communication and faster remediation. Ensure the DSPM goes beyond raising alerts and helps you understand and effectively communicate violations to data owners.

MULTIPLE FACTORS FOR RISK PRIORITIZATION

A DSPM solution should prioritize exposures for remediation based on a risk profile that takes into account multiple factors, such as data sensitivity level, data volume, data exposure, and data security posture. Sensitivity types should be customizable and then associated with the proper risk profile.

GUIDED REMEDIATION FOR POLICY VIOLATIONS

In addition to identifying policy violations, a DSPM solution should provide guided remediation to address them. This should include guidance on critical data security practices, such as restricting third-party access to sensitive data, as well as data hygiene best practices, like deleting redundant or obsolete data. To ensure user-friendliness, the best DSPMs accept data-centric policies that work with various technologies throughout the organization and provide technology-specific remediation suggestions.

When a DSPM detects a policy violation, it should provide granular context about the policy, describe how the violation occurred, and offer technology-specific actionable next steps for remediation. Moreover, to ensure efficient remediation of policy violations, the DSPM should prioritize user-friendliness and provide streamlined remediation options, such as escalation, reassignment, and auto-remediation. To enable these options, the DSPM should go beyond read-only privileges and have the ability to configure parts of your environment parameters. Therefore, it's essential to select a DSPM that offers options for both a read-only installation and an action-enabled one, with ready integrations to other workflow platforms (see more on this in Criteria 3 below).

EXAMPLES OF DATA-CENTRIC POLICY VIOLATIONS

VIOLATION CATEGORY	VIOLATION EXAMPLE	POLICY FOR REMEDIATING AND PREVENTING VIOLATION
Overexposed Data	Sensitive data is inadvertently made public and exposed to the wrong users	Ensuring proper governance of that data and ensuring it only resides where authorized user access applies.
Unprotected Data	Sensitive data is not actively protected by applicable means such as encryption, retention policy etc.	Enabling data security best practices such as encryption, masking, retention period and activity logging.
Misplaced Data	Sensitive, sovereign data is in forbidden geo-location	Removing protected data from unauthorized locations, then automatically apply residency rules to sensitive assets in the future.
Redundant Data	A team member copied a sensitive asset then never used it again. It's now sitting somewhere, unknown and unmanaged.	Automatically removing all sensitive assets that weren't used in X days

Built-in data security policies provide a proactive approach to security, privacy, and compliance that supports innovation rather than hindering it. The best DSPMs take a multi-faceted approach by identifying policy violations upon deployment, prioritizing them based on risk, and continuously detecting and alerting on violations in near real-time as data proliferates.

This process fosters a culture of data risk management and data hygiene throughout your organization without compromising agility. These policies reinforce compliance and reduce the risk of data breaches and other security incidents.

CRITERIA 3 – OPERATIONALIZATION FOR YOUR WHOLE ORGANIZATION

To make automated data discovery, classification, and policy frameworks a reality, operationalizing the DSPM solution is crucial. With today's data democratization and self-service cloud infrastructure, data responsibility is spread across multiple organizational units, including data security, privacy teams, development, DevOps, and business users. This means that your security team won't be the only group interacting with your chosen DSPM solution.

To ensure that a given DSPM is an excellent, interoperable fit for your entire organization, consider the following features:

ZERO DISRUPTION ARCHITECTURE AND SEAMLESS SYSTEM INTEGRATIONS

When searching for a DSPM solution, it's important to look for one that is easy to deploy, integrates seamlessly with your existing workflows, and can integrate with CSPMs, enterprise data catalogs, ITSM and SIEM tools, team communication services, and privacy and governance management tools via APIs. The platform should also support multiple CSPs and data warehouses like AWS, Azure, GCP, Snowflake, and Databricks. Additionally, it should be a zero-disruption platform, which means it must be agentless and asynchronous to prevent any disruption to your cloud environment.

INSTALLATION SIMPLICITY AND FAST TIME-TO-VALUE

The installation process should be simple, with a short deployment time and no connector configuration or user/asset credentials needed. The DSPM solution should deliver significant results within a week of implementation, revealing a complete set of detailed findings of sensitive, regulated, and proprietary data throughout your organization.

SELF-EXPLANATORY FEATURES AND ACCESS TO SUPPORT

The DSPM solution should have a modern, easy-to-use interface with self-explanatory features, providing a wide variety of self-service resources and technical documentation. In case you have specific or highly-technical questions, the vendor should provide expert technical support with quick response time and expertise.

UNIFIED CONSOLE

The DSPM solution should include centralized reporting and drill-down views based on governance, privacy, or security functions. You should be able to view the entire organization by business unit, environment, and asset, then zoom in to understand the sensitivity level, ownership, volume, usage, and purpose of each data finding.

RUBRIK'S SOLUTION FOR DSPM

Rubrik DSPM provides organizations with the data visibility and control they need to reduce the risk of data exfiltration and minimize the impact of cyberattacks. This includes known and unknown—structured, semi-structured, and unstructured—sensitive data across on-premises, cloud (AWS, Azure, GCP), data warehouse (Snowflake, BigQuery), and SaaS environments

For cloud environments, Rubrik DSPM takes an API-only approach, without any agents and without removing sensitive data, thereby avoiding regulatory compliance issues. Rubrik DSPM is embedded within an organization's cloud accounts and analyzes only metadata so source data never leaves the cloud.

Rubrik DSPM analyzes access, usage patterns, and security posture, and provides actionable, guided remediation for data security risks.

If you want to learn more about Rubrik DSPM, contact us today for a demo!


rubrik