BigID

The State of Data Security

Top Data Security Challenges and Priorities

Table of Contents

Overview & Objective

Security leaders rank finding and eliminating dark data as the data security initiative of greatest concern for their organizations to address.

Data is every organization's single most valuable asset - an asset dependent on to drive critical decisions every day. Much of this data is highly sensitive or critical, and in some cases vulnerable to accidental exposure or malicious activity. With the accelerating growth of public, private, hybrid, and multi-cloud models, sensitive or critical data proliferates throughout data environments at unbelievable rates. As the footprint of this sensitive or critical data expands, so does organizational risk.

BigID set out to obtain better insight into the state of the data security practice in today's organizations. This report dives deep into the individual limitations their IT and Security teams face and the resulting issues that are born out of current tools, methods, and processes that they are leveraging.

We surveyed over 400 enterprise technology leaders to explore how they currently manage and protect sensitive and critical data, the challenges they face when doing so, and how they can improve their data security strategy.

Executive Summary

- Most organizations are currently concerned with finding and managing unstructured and dark data across their environment.
- Organizations continue to struggle to confidently enforce data policies, remediate data, and leverage automation in day-to-day security operations.
- The vast majority of organizations fail to confidently find and classify all of their sensitive or critical data - and believe it's equally challenging in the cloud as it is on prem.
- Almost every organization sees synergistic value when security and privacy teams share a single source of truth about their data.
- Interoperability continues to be a critical factor for most when considering future security investments.
- More than a quarter of organizations think their DLP tools fail to fulfill their data protection needs.

Section: 1

Data Landscape: The Night Is Dark and Unstructured

Data Landscape: The Night Is Dark and Unstructured

Dark data is data that organizations are unaware about, but typically make up over half of all data in existence and can be highly sensitive or critical in nature - examples being credit card information, intellectual property, financial data, and highly regulated data. This survey found that the vast majority of Security and IT leaders (84%) are concerned about finding and eliminating dark data. The majority of dark data is unstructured and likely to be proliferating across the data ecosystem at a rapid rate.

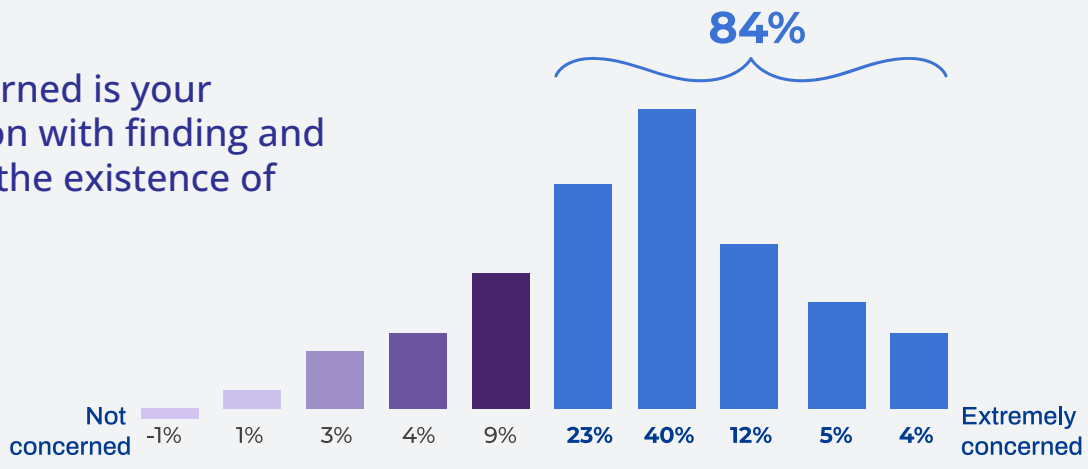
Given dark data's expanding footprint and its unstructured nature, finding and securing it with today's tools and methods can be challenging.

Traditional data discovery tools lack the ability to find and manage unstructured data because they utilize limited, pattern-matching-based detection capabilities that tend to be exclusively focused on Regular Expressions (RegEx), and are often limited in data source coverage.

84% of organizations are concerned about finding and eliminating dark data.

This leaves gaps in identifying potentially sensitive, critical, and regulated data - especially since you can't protect what you can't see. As the amount of dark data increases, so does its organizational risk — especially if this data is sensitive or critical. It's more important than ever for organizations to be able to automatically identify their dark data across their entire data landscape: from on-prem to cloud, including both data at rest and data in motion. It's critical for Security and Risk leaders to implement data discovery solutions that leverage advanced classification techniques using AI and ML to see all types of data - structured, unstructured, and semi-structured - so that they can mitigate their dark data with certainty.

Q How concerned is your organization with finding and mitigating the existence of dark data?

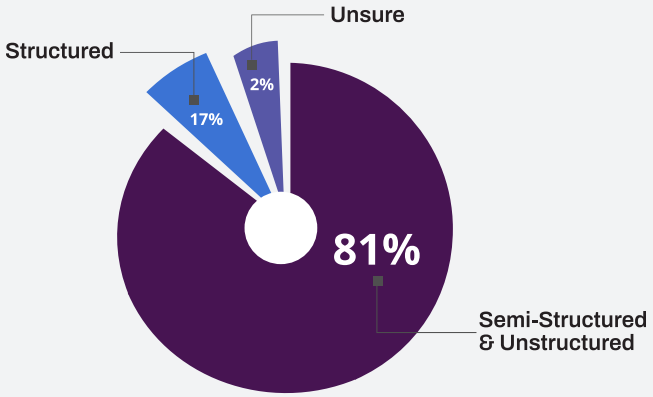


More than 8 out of 10 organizations consider unstructured and semi-structured data the most difficult type of data to locate and manage.

By 2025, IDC estimates that there will be 163 zettabytes of data in the world, 80% of which will be unstructured. Unstructured data is data that is not structured in common database formats.

This survey found that over eight out of 10 Security and Risk professionals consider unstructured data the most difficult type of data to locate and manage. And second to that? Semi-structured.

Q What type of data has been the most difficult to locate and manage?



Many organizations struggle with data discovery and insight, especially within unstructured data - including files, documents, spreadsheets, text, and more. Even more concerning? The growth in the intelligent edge and 5G networks has led to an explosion of mobile, IoT, and OT machines generating an immense amount of unstructured data. The difficulty in protecting this data lies in the complexity and sophistication of this data when trying to discover and classify the nature of its sensitivity or criticality.

Unstructured data traditionally contains multitudes of sensitive information (just waiting to be breached) - but also traditionally takes a long time to scan in the first place.

as regular expression (RegEx), keywords, and pattern matching. These outdated approaches are limited in their ability to completely discover and classify unstructured data - leaving not only broad blind spots in the attack surface, but creating noise - leading to high rates of false positives and inaccurate data classification.

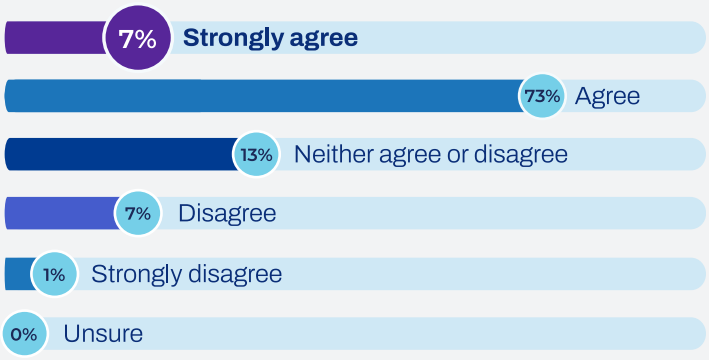
Organizations that are prioritizing unstructured data protection need to leverage advanced machine learning (ML) and artificial intelligence (AI) to augment traditional methods - and solve for the complexity around working with unstructured data. Similar to the dark data problem, organizations must adopt data discovery solutions that offer the ability to see all types of data, everywhere, regardless of location.

Section: 2

Taking Action: Visibility Without Control Limits A Strong Security Posture

Taking Action: Visibility Without Control Limits A Strong Security Posture

Q To what extent do you agree with the following statement: "My organization is properly enforcing policy around sensitive data for controls".



Organizations typically resort to a "spray and pray" approach to policy management, crossing their fingers that their entire estate is covered. However, this blanket strategy proves difficult since many teams lack complete data visibility and control due to the limitations of their existing, mostly legacy, toolsets.

Even worse, they also fail to verify that every part of their data ecosystem is being covered. Security and Risk leaders can believe they are properly enforcing data protection policy, but without the proper tools to verify policy application, they're left in the dark.

Over 50% of organizations rate their ability to remediate data and automate tasks as unsatisfactory.

Only 7% of organizations strongly agree that they can enforce data protection policies effectively.

hey often lack the certainty gained through reporting and auditing capabilities and the assurance that their most sensitive or critical data is being protected everywhere, at all times. The lack of certainty hinders their ability to effectively deploy policies for proper data control.

How can Security and Risk professionals feel more confident in their ability to enforce data protection policies? They need to be able to extend their policies across all types of data, everywhere, regardless of where it exists whether it be in the cloud or on prem. Along with coverage, organizations must be able to apply these policies across the right types of data, consistently and verify that these policies are working. Coverage, context, consistency, and certainty are critical factors for effective data protection policy programs.

Carrying out remediation measures quickly and effectively to prevent and mitigate the damage of a possible security incident is a critical component of any data security strategy. Surprisingly, this report finds that over half of organizations consider their ability to remediate data unsatisfactory.

Remediation starts with finding and understanding data to create a current and complete data asset inventory. With visibility comes the ability to carry out actions and corrective measures on your data — whether it needs to be minimized, quarantined, deleted, masked, or have some other action taken on it. IT and Security teams need to be able to streamline the remediation workflow by assigning remedial actions to the right people so that they can take the right protective measures on the right data. Unfortunately, today's fragmented data security stack lacks a seamless way to bridge the gap between insight and action. Data security tools lack proper coverage across the environment, interoperability with other tools to enable the right actions, and scale to support high volumes of data.

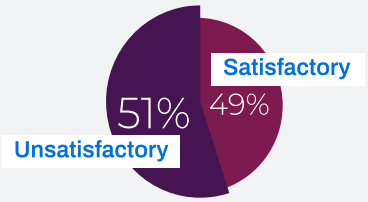
Most security organizations are resource-constrained and continually looking for ways to drive efficiencies and reduce complexity as means to become more agile. Security practitioners and incident response teams need to keep pace with the speed of adversarial threats. Most organizations deploy a combination of legacy point tools for data management and security.

This patchwork of tools is brittle and prone to failure, and, worst of all, lacks end-to-end automation and speed teams need to get the job done. This is likely why the report found that over half of organizations consider their ability to automate tasks as unsatisfactory. Legacy tools like Data Loss Prevention (DLP) require manual deployment, configuration, tuning, and maintenance, which creates inefficiency and a higher risk of human error.

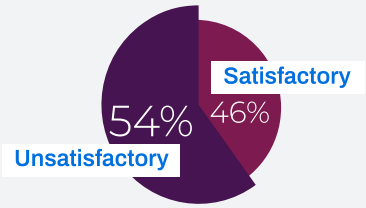
Traditional data discovery and classification tools also lack AI and ML capabilities that greatly help security teams be more accurate and mitigate the number of noisy false positives that continue to drive alert fatigue.

Organizations need the flexibility to remediate their data the way they want, leveraging as much automation in the process as possible. Security teams need data security solutions that enable orchestration through remediation workflows in a way that makes sense for their organization. This enables organizations to get the right people to take the right actions on the right data, seamlessly. Security teams are best positioned to automate manual tasks when they implement data security methods that incorporate AI and ML with respect to finding and controlling their data. AI and ML aren't luxuries - they are table stakes necessities to completely and confidently find and control your data and keep pace with the agility of adversarial threats.

Q Rate the effectiveness of your CURRENT data security tools, methods, and practices across the following capabilities.



Taking remedial actions



Automating manual tasks

Section: 3

Data Discovery: Security Starts with Knowing What You Have

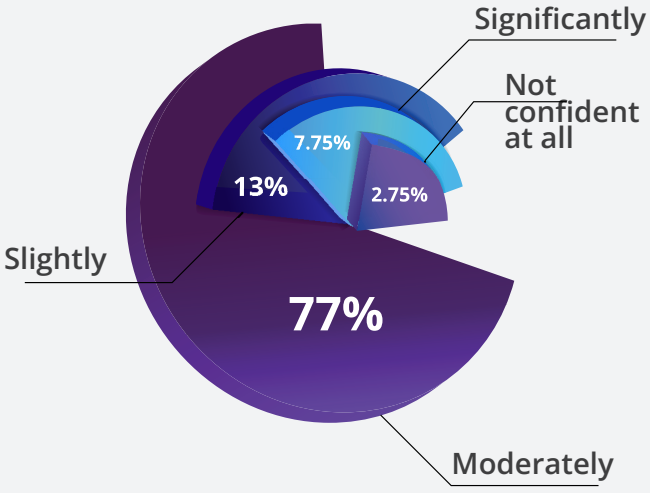
Data Discovery: Security Starts with Knowing What You Have

Only 8% of organizations can confidently find ALL their sensitive or critical data across their environment.

Organizations lack the complete ability to discover, understand, and classify all their data across their environments - only 8% of the respondents were confident in their ability to discover all sensitive and critical data.

Data discovery is a foundational component of any strong data security practice. Organizations must be able to confidently find and manage information (whether sensitive, critical, or not). Data discovery is arguably step zero in the process of protecting your data. Teams must leverage data discovery solutions that offer complete visibility and coverage - across all file types and data environments. You can't protect what you can't see, and therein lies the problem.

Q How confident are you in your organization's ability to discover all sensitive or critical data across your environment?



Most data discovery solutions today exclusively focus on structured or unstructured data — but rarely both. Unstructured data, such as rich media and text files, is difficult to find and ascertain, requiring more intelligent tooling and analysis. Furthermore, traditional data discovery solutions work primarily well in the cloud but struggle to completely and accurately support on-prem environments. This poses the greatest challenge for organizations that deploy primarily hybrid or on-prem environments.

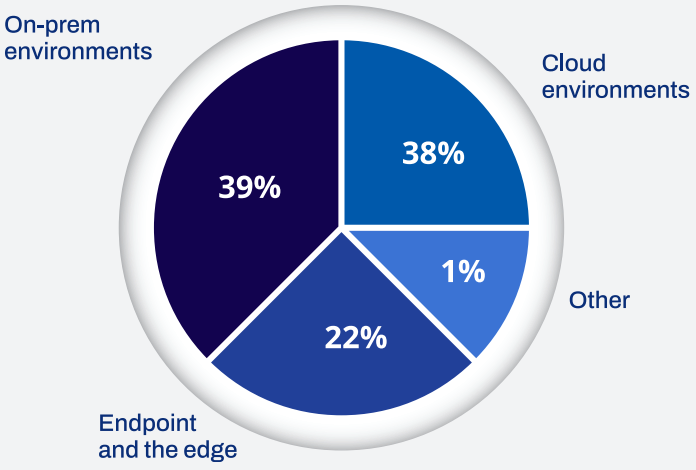
Almost all organizations consider discovering data across the cloud as equally challenging as across on-prem environments.

While cloud adoption shows no signs of slowing down, the majority of large enterprise organizations have yet to fully realize the potential of the cloud. Over the past few decades, many of these organizations have built up a mammoth footprint of non-cloud-based data sources that aren't deprecating anytime soon.

Most organizations that are currently in their cloud migration journey optimize by leveraging a hybrid environment in the interim, as they gradually scale down their on-prem deployment model over the long-term. As a result, enabling complete coverage across any environment — cloud, on-prem, hybrid — is critical for any robust data discovery and security practice. Unfortunately, most traditional data discovery solutions today fail to properly support on-prem discovery, as they do for the cloud. Almost every respondent in the report considered discovery data across the cloud as equally challenging as across on prem.

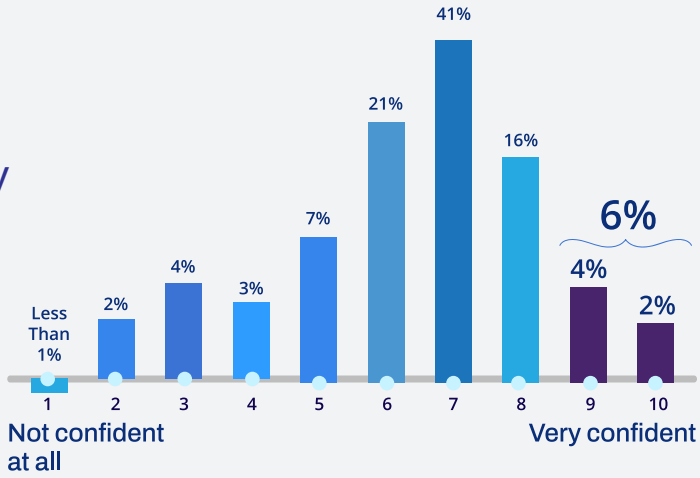
On-prem environments are bespoke and discovery can be challenging. Proper data protection must extend across any type of environment and across all types of data in order for Security teams to confidently safeguard every critical data asset. Security teams can alleviate their concerns by leveraging a federated, data discovery platform that supports any data environment, regardless of where it exists - in the cloud, on prem, or both - and all data types and sources. This flexibility gives teams the peace of mind knowing that their entire data estate is covered, irrespective of their environment now or into the future. A federated data discovery becomes the single source of truth about your data - enabling IT and Security teams to find and manage all of their data through a single pane of glass.

Q In which part of the network do you find it most challenging to discover sensitive or critical data?



Only 6% of organizations are very confident in their ability to classify their data by its sensitivity or criticality.

Q Rate your organization's ability to understand/classify this data by sensitivity or criticality?



Understanding the context behind all your data is just as important as knowing where it is located. It's no surprise that less than a quarter of respondents can confidently classify the data they discovered with respect to its sensitivity or criticality. As mentioned previously, unstructured data is difficult to manage and classify due its inherent complexity. Traditional data classification solutions leverage simple pattern-based matching like regular expression (RegEx).

Even more difficult? Classification requirements can be unique to an organization and vary in sophistication. For example, healthcare organizations may employ a set of sensitivity or criticality definitions requiring special classifiers, while a retail organization may deploy a different set. Consistently classifying data according to an organization's own definition of sensitivity or criticality can be extremely difficult across disparate data environments that leverage their own classification frameworks. Multi- and hybrid-cloud environments contain a variety of data sources, each with their own classification or data labeling frameworks. The lack of conformity across sensitivity or criticality definitions fails to provide Security teams with a confident classification process.

IT and Security teams can solve their challenges by adopting data discovery and security solutions that leverage modern classification techniques. Modern data classification methods go beyond ReGex classifiers and are augmented with artificial intelligence (AI) and machine learning (ML) so that teams can fully gain contextual insight behind their data. These advanced classification techniques cover what traditional classification techniques miss and to improve over time, learning the nuances of sensitive or critical data unique to an organization.

Section: 4

Security Through Synergy

Security Through Synergy

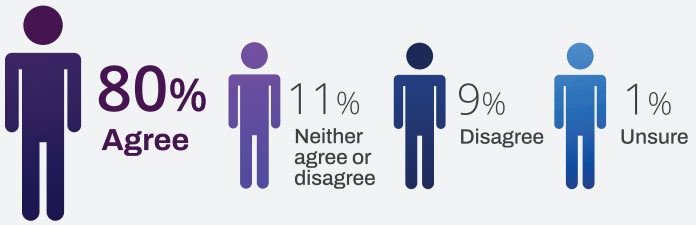
80% of organizations agree that it is critical for Security and Privacy teams to share a current and complete data asset inventory to drive decision-making.

Historically, data security and data privacy have been largely viewed as two distinct practices, each consisting of a set of workflows and tools unique to their own divisions. Today, this is becoming less true. Data Security teams are responsible for implementing protective controls around all of their sensitive or critical data, while data privacy teams are largely concerned with adhering to individual or consumer rights around that data. In order for both teams to successfully meet their objectives, they need to obtain a solid understanding of all of their data, anywhere it exists. It is why the report found that the majority of organizations (80%) strongly believe sharing a current and complete inventory of data between Security and Privacy teams is critically important to drive the right decisions.

This starts with building a current and complete data asset inventory that serves at the bedrock for both security and privacy initiatives. Security teams can't protect what they can't see, nor can privacy teams meet regulations without knowing what user data they have. Both teams largely benefit from sharing a current, complete, single source of truth about the state of all data assets.

The convergence between Security and Privacy teams is best exemplified through recent National Institutes of Standards & Technology (NIST) frameworks. NIST frameworks serve as baseline, risk-based guidance for better risk management and reduction across the IT estate. Both the security and privacy frameworks share common workflows that interlap on data discovery, protection, and response.

Q To what extent do you agree with the following statement: "Both our security and privacy teams share a common, current, and complete sensitive/critical data inventory to drive their decision-making."



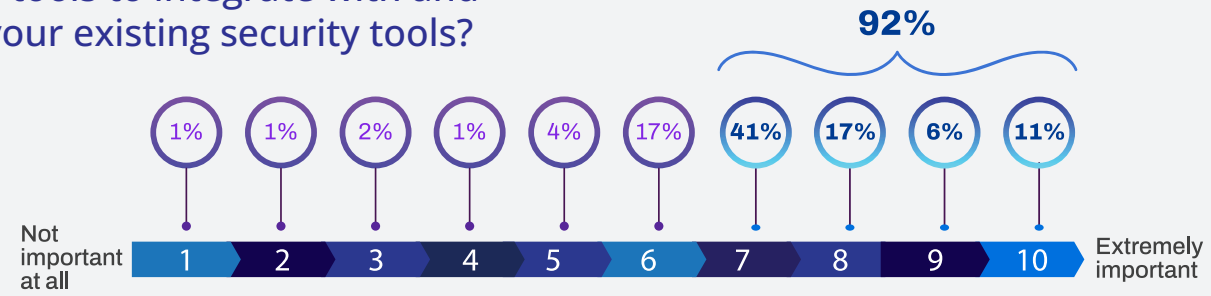
As the needs of Security and Privacy teams begin to converge, it's imperative for organizations to support the merger through a shared set of tools and processes that both can rely on. There is no better example of this than a federated data security and privacy platform solution that answers the needs for both organizations - sensitive or critical data discovery, access intelligence, and remediation control to meet privacy and security initiatives. A converged platform provides both teams a current, common, and complete picture of their most important data. This shared source of truth drives synergistic behavior and eliminates the needs for multiple point solutions - saving organizations valuable time and money in the process.

A majority 92% of organizations agree that interoperability is important when considering future security investments.

When considering future security vendors, most large enterprise organizations prefer to augment existing tools rather than displace them due to strong technical debt and expertise with incumbent technologies. Per the report, the vast majority of organizations (92%) believe interoperability is critically important when considering future security purchasing decisions.

Augmenting existing tools mitigates the possibility of a possible IT disruption a displacement could potentially cause while accelerating the onboarding of new technologies. Organizations prefer new vendors to enrich and enhance their existing security stack wherever possible through seamless, API-led integrations. Layering new technologies on top of legacy solutions may alleviate any perceived technical gaps while also extracting more value out of existing investments. Therefore, it is critical for new security technologies to integrate with and enrich popular tools used within the security stack today at most large enterprise organizations.

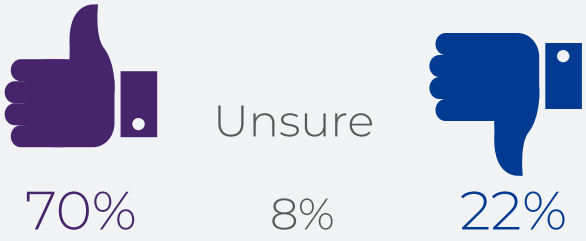
Q How important is the ability for future security tools to integrate with and enrich your existing security tools?



Almost 1 in 4 organizations feel their existing DLP tools are inadequately fulfilling their data protection needs.

Data loss prevention (DLP) was the industry's first answer to the challenge of protecting an organization's sensitive or critical data. However, like most legacy solutions, DLP has not evolved fast enough to keep pace with the environments of today, much less ones of tomorrow. This is why a significant number of organizations (25%) feel that DLP tools inadequately fulfill their data protection needs.

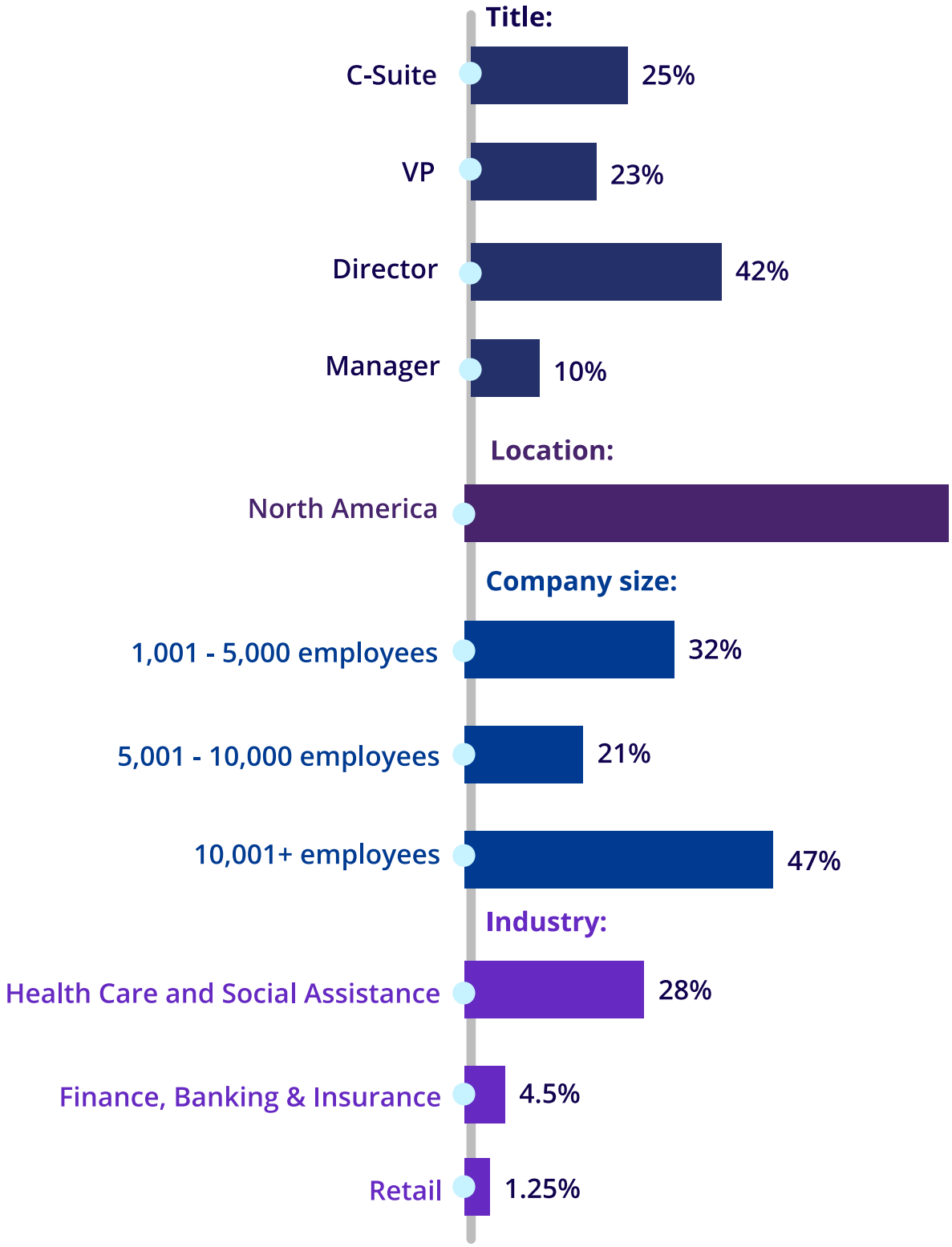
Q Thumbs up/down/?: Are your existing data loss prevention (DLP) tools sufficient enough to fulfill your data protection strategy?



These tools were primarily designed for on-prem environments where data was protected simply through a security control point or channel. Legacy tools like DLP rely on pattern-based matching which limits classification abilities and cannot always scan both structured and unstructured data. Many DLP tools also require manual deployment, configuration, tuning, and maintenance, which creates excess overhead and increases the risk of human error. On top of this, they are somewhat inaccurate and noisy, requiring multiple policies that can't keep up with the fluidity of data which inevitably generates too many false positives for Security teams.

DLP has its place within an organization's security stack, but by itself fails to provide the complete protection that Security teams should expect. Instead, teams must augment DLP with solutions that enhance insight, context, and reach for better data controls. Supplement DLP with advanced classification methods that can better determine sensitivity and criticality across all types of data - structured, unstructured, semi-structured, etc. Combine advanced classifications with solutions that also offer the ability to consistently tag and label data across the entire estate - not just a subset - so that you can enrich and extend policies set through DLP everywhere your data lives.

Respondent breakdown



Contact

Neil Patel
BigID
641 Avenue of the Americas, 5th Fl,
New York, NY 10011
npatel@bigid.com