

[← All articles](#)

Sensitive Data Discovery



Wiz Experts Team

October 31, 2024

5 minute read

[Download Cheat Sheet](#)[Get a Demo](#)

What is sensitive data discovery?

Sensitive data discovery and classification identifies and categorizes confidential information—like personally identifiable information (PII), protected health information (PHI), financial data, and intellectual property—that's stored in databases, documents, emails, and elsewhere to ensure compliance and protect sensitive data.

Every business handles sensitive information of different kinds from financial data (like payroll information) to customer information (including addresses and social security numbers). If you don't protect this information, you risk financial loss or reputational damage. And if attackers gain access to confidential data, they could use it for ransom or wreak further damage, like taking down mission-critical servers.

Table of contents

[What is sensitive data discovery?](#)[Why not secure everything?](#)[How to discover, classify, and secure your sensitive data](#)[Sensitive data discovery use cases](#)[Adopting DSPM for sensitive data discovery and classification](#)[Wiz DSPM makes it all possible](#)

GenAI Security Best Practices [Cheat Sheet]

Discover the 7 essential strategies for securing your generative AI applications with our comprehensive GenAI Security Best Practices Cheat Sheet.

[Download Cheat Sheet >](#)

See Wiz in Action

Start your interactive tour and see how Wiz can secure your cloud from code to runtime.

[Start product tour](#)

Securing your data begins with two principles:

- **PRINCIPLE #1: You can't protect what you don't know about**

Data discovery identifies all the places your data is hiding. That may be on-premises, the cloud, emails, productivity suites, databases, and more.

- **PRINCIPLE #2: To protect your treasures, you have to find the gold**

Data classification sorts your data, assigning labels to it based on its sensitivity.



Orca Security Ltd.
Exhibit 2157
Wiz v. Orca
IPR2024-00865
Ex. 2157-001

- **PRINCIPLE #2: To protect your treasures, you have to find the gold**

Data classification sorts your data, assigning labels to it based on its sensitivity.

Finding and protecting data in the cloud is particularly difficult because of shadow data, [meaning unrecognized or unmanaged data in the cloud](#).

Shadow data is created because of the new ways of working in the cloud. Data is scattered across multiple locations, making it hard to track. Cloud providers also have separate security policies, which makes tracking and policy enforcement more difficult. Cloud environments are also constantly changing, so visibility is another huge challenge.

In this post, we'll find out why the sensitive data discovery process is so important—along with some of the main challenges. We'll see how companies tackle the daunting task of classifying their data. Then we'll explore expert approaches and strategies to make it simpler.

Why not secure everything?

As you've probably guessed, sensitive data discovery and classification can be a massive and tedious task. So why not just protect all your data with the highest level of security?

One reason is that higher security comes at a higher cost, and it's more complex to manage. Another reason? Excessive controls actually make your business less efficient. Imagine needing full multifactor authentication—like verifying on your phone—just to access the agenda for a team meeting. The file would be secure, but staff would be very irritated. They might not even bother checking the agenda if it's too difficult. Finally, protecting everything means you're not focusing on the assets and resources that need the most protection.

Not all your data has equal value (that hilarious meme on the office WhatsApp compared to the source code for your biggest product, for example). Teams must focus on the most critical assets—and data discovery and classification help you do exactly that.

How to discover, classify, and secure your sensitive data

As we saw above, you can't apply the same level of security to all your data. So you need to start by classifying your data—finding the crown jewels that really need top-level protection [from today's leading threats](#).

Which classification levels should you use?

Classification levels create common language across your organization. This lets teams and departments decide which type of data belongs to which category:

Classification level	Definition	Examples
Public	Data that can be publicly	Your website, marketing materials, contact lists for sales, support,

Table of contents

What is sensitive data discovery?

Why not secure everything?

How to discover, classify, and secure your sensitive data

Sensitive data discovery use cases

Adopting DSPM for sensitive data discovery and classification

Wiz DSPM makes it all possible

See Wiz in Action

Start your interactive tour and see how Wiz can secure your cloud from code to runtime.

[Start product tour](#)



Classification level	Definition	Examples
Public	Data that can be publicly shared	Your website, marketing materials, contact lists for sales, support, and other departments, along with a support wiki for users.
Internal only	Data that isn't intended for the public	Organizational charts and internal sales materials like playbooks and battlecards aren't exactly secret, but they are reserved for internal use.
Confidential	Data that could harm your business if released to the public	This includes vendor contracts, employee applications and reviews, along with some employee and customer data.
Restricted	Data that could severely harm your organization if released	Includes intellectual property, credit card information, social security numbers, and health information. This level of data is generally governed by regulations (e.g., HIPAA and PCI DSS).

Table of contents

[What is sensitive data discovery?](#)

Why not secure everything?

[How to discover, classify, and secure your sensitive data](#)[Sensitive data discovery use cases](#)[Adopting DSPM for sensitive data discovery and classification](#)[Wiz DSPM makes it all possible](#)

See Wiz in Action

Start your interactive tour and see how Wiz can secure your cloud from code to runtime.

[Start product tour](#)

Which classification approaches are available?

After determining the classification level (the four categories above), it's time to pick a classification approach:

Classification approach	Definition	How it works
Content-based classification	Analyzes files for sensitive information like PII, financial data, medical information, or intellectual property	Automated: Uses techniques like regular expressions, pattern recognition, and fingerprinting to identify specific data types in file contents.
Context-based classification	Uses information about files (as opposed to content) to determine what level of protection is needed	Automated: Considers metadata, file location, access patterns, user roles, and sharing settings to assess data sensitivity based on contextual factors.
User-based (or manual) classification	Flags sensitive documents during creation, editing, or review	Manual: Performed by users who label data according to its sensitivity level, often using guidelines established by security policies.

Which classification strategy is best?

Manual classification is the most time-consuming but may make it easier to identify sensitive content. Manual classification also adapts easily to changing needs and a variety of sensitive data types. On the other hand, it does open you up to insider threats.

Data discovery and classification tools automatically classify data based on content and context, eliminating the need for tedious manual work. Automated classification is fastest, and it scales to effectively handle cloud-based data. But there's a greater risk of false positives. [AI-based security solutions](#) can bring more nuance to automation, lowering the risk of false positives.



[based security solutions](#) can bring more nuance to automation, lowering the risk of false positives.

Sensitive data discovery use cases

Healthcare providers

Healthcare organizations like hospitals and clinics handle extensive patient health information (PHI) and must [secure medical records data to comply with HIPAA regulations](#). They also need to protect clinical trial data to ensure research integrity and patient privacy, and in some cases, they are subject to other regional laws or privacy regulations.

Financial services

Finance companies such as banks and insurance companies have to protect customer financial records to comply with privacy regulations. They also need to protect any proprietary investment algorithms (intellectual property), as these give them a competitive advantage.

Government agencies

A federal agency handling taxation, for example, must secure citizen income data and financial reports to meet data protection regulations. Government agencies such as defense ministries or FEMA in the U.S. must also protect classified documents for regional and national security.

Software vendors

Software vendors must [keep customer data safe to comply with data privacy regulations](#). They also store mission-critical R&D and IP such as source code and patentable algorithms that require protection.

Adopting DSPM for sensitive data discovery and classification

Data security posture management (DSPM) is one of the best sensitive data discovery tools, letting you continuously manage and secure sensitive data across multiple cloud environments.

DSPM gives you visibility into data security risks and compliance. With customizable rules, it automates discovery and classification. DSPM then provides ongoing monitoring, alerting, and policy enforcement. And with clear, meaningful reporting on your data security posture from DSPM, regulatory compliance becomes much simpler.

Table of contents

What is sensitive data discovery?

Why not secure everything?

How to discover, classify, and secure your sensitive data

Sensitive data discovery use cases

Healthcare providers

Financial services

Government agencies

Software vendors

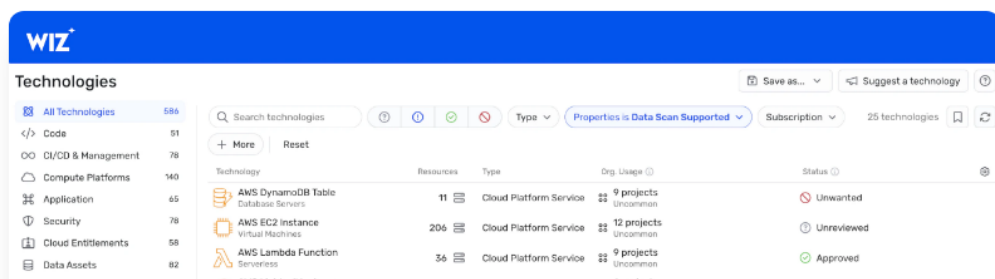
Adopting DSPM for sensitive data discovery and classification

Wiz DSPM makes it all possible

See Wiz in Action

Start your interactive tour and see how Wiz can secure your cloud from code to runtime.

Start product tour



Technology	Resources	Type	Org. Usage	Status
AWS DynamoDB Table Database Servers	11	Cloud Platform Service	9 projects Uncommon	Unwanted
AWS EC2 Instance Virtual Machines	206	Cloud Platform Service	12 projects Uncommon	Unreviewed
AWS Lambda Function Serverless	36	Cloud Platform Service	9 projects Uncommon	Approved
AWS Lightsail Instance Virtual Machines	1	Cloud Platform Service	6 projects Uncommon	Unreviewed
AWS RDS Aurora MySQL Cluster Database Servers	3	Cloud Platform Service	9 projects Uncommon	Unwanted
MariaDB Server Database Servers	5	Server Application	13 projects Uncommon	Unreviewed
MongoDB Database Servers	13	Server Application	19 projects Common	Approved
MySQL Database Servers	7	Server Application	15 projects Common	Unreviewed
OCI Bucket Storage Buckets	38	Cloud Platform Service	6 projects Uncommon	Unreviewed
OCI Compute Instance Virtual Machines	60	Cloud Platform Service	6 projects Uncommon	Unreviewed
PostgreSQL Database Servers	8	Server Application	13 projects Uncommon	Approved
Snowflake Database Database Servers	14	Cloud Platform Service	1 project Rare	Unreviewed
Snowflake Schema Database Servers	159	Cloud Platform Service	1 project Rare	Unreviewed

Figure 1: Wiz's DMSM solution includes data lineage mapping so you can see the entire data lifecycle at a glance

TL;DR? DSPM helps you tackle the biggest challenges of sensitive data discovery and classification. That includes finding, identifying, and securing all your cloud-based data, including shadow data.

Wiz DSPM makes it all possible

Data isn't your only security concern. In the real world, your security teams have a lot to keep track of. They have to monitor network access, keep malware in check, and track vulnerabilities. And siloed data security tools—including separate DSPM tools—fragment visibility and increase complexity for your entire team.

A **cloud native application protection platform (CNAPP)** like Wiz, on the other hand, combines all your core security tasks behind a single pane of glass, including:

- **Cloud security posture management (CSPM)**: Flags misconfigurations to help secure cloud environments
- **Cloud workload protection (CWP)**: Monitors and protects cloud workloads, including containers and serverless
- **Cloud infrastructure entitlement management (CIEM)**: Manages user access and privileges connected with cloud resources

With integrated DSPM, Wiz takes a load off your mind with a streamlined approach to data security. It finds and classifies data across all your cloud environments, including PaaS, IaaS, and DBaaS, cutting your risk of shadow data. Wiz lets you enforce data security controls and policies uniformly, regardless of how and where your data is stored.

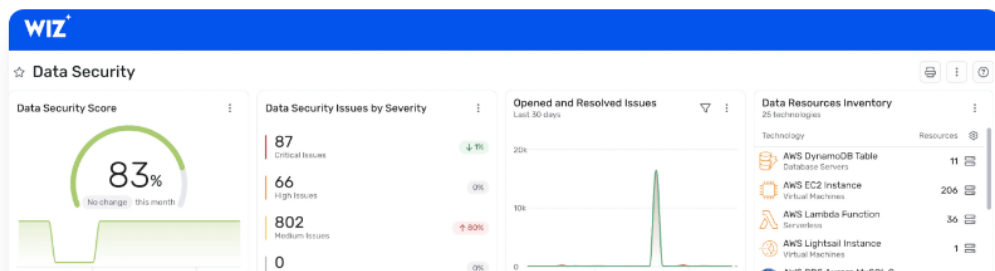


Table of contents

What is sensitive data discovery?

Why not secure everything?

How to discover, classify, and secure your sensitive data

Sensitive data discovery use cases

Healthcare providers

Financial services

Government agencies

Software vendors

Adopting DSPM for sensitive data discovery and classification

Wiz DSPM makes it all possible

See Wiz in Action

Start your interactive tour and see how Wiz can secure your cloud from code to runtime.

[Start product tour](#)

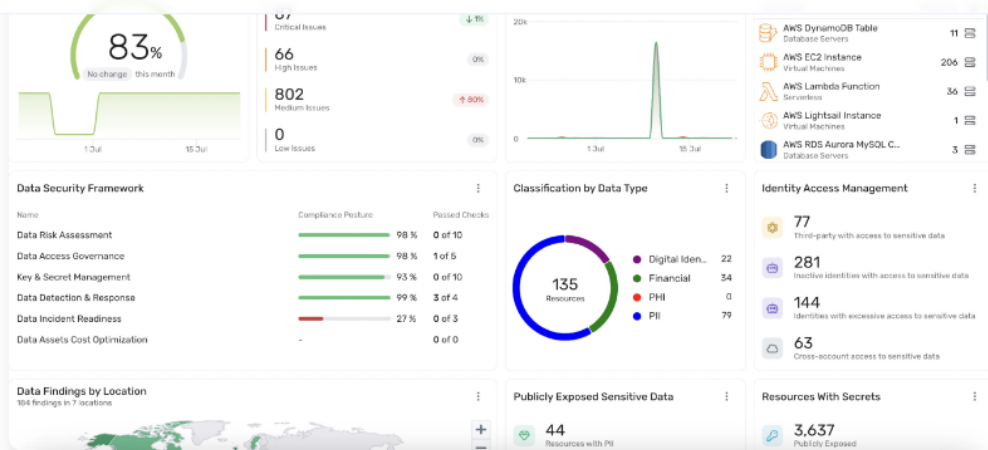


Figure 2: Wiz's all-in-one dashboard gives you uniform control over all your cloud environments

With real-time monitoring, Wiz correlates data risks with other cloud risks, uncovering threats before they impact your organization. That includes sensitive training data in AI pipelines—one of today's fastest-growing data risks.

Most importantly, Wiz is easy to set up and roll out, with agentless scanning for total security coverage without blind spots.

Ready to find out how integrated DSPM can protect your sensitive data? Get a [personalized demo](#) to see how Wiz can help you reduce your attack surface, meet compliance standards, and secure your complex multi-cloud development and data environments.

Table of contents

What is sensitive data discovery?

Why not secure everything?

How to discover, classify, and secure your sensitive data

Sensitive data discovery use cases

- Healthcare providers
- Financial services
- Government agencies
- Software vendors

Adopting DSPM for sensitive data discovery and classification

Wiz DSPM makes it all possible

See Wiz in Action

Start your interactive tour and see how Wiz can secure your cloud from code to runtime.

[Start product tour](#)

Protect your most critical cloud data

Learn why CISOs at the fastest companies choose Wiz to secure their cloud environments.

[Get a demo >](#)



Explore more on this topic

Cloud Security
Fundamentals

Vulnerability and Risk
Management

Identity Security

- What is CIEM? Cloud Infrastructure



Cloud Security Fundamentals

- [Cloud Security 101 for CloudSec, AppSec, and SecOps Managers >](#)
- [Cloud Security Strategy >](#)
- [What is Cloud Security Architecture? Principles, Framework, and Architecture Assessment >](#)
- [What is CNAPP? \[Cloud-Native Application Protection Platform\] >](#)
- [What is CSPM? >](#)

Vulnerability and Risk Management

- [What is Vulnerability Management? Steps and Components to Implement Vulnerability Management >](#)
- [Vulnerability Scanning: How it Works, Why it Matters, and What You'll Find >](#)
- [Vulnerability Management Best Practices >](#)
- [Vulnerability Prioritization in the Cloud: Strategies + Steps >](#)
- [Risk-Based Vulnerability Management >](#)
- [Top OSS vulnerability management tools >](#)

Identity Security

- [What is CIEM? Cloud Infrastructure Entitlement Management Use-Cases, Challenges, and Benefits >](#)
- [What is IAM Security? Components, Features, Best Practices >](#)
- [Effective Permissions: A Security Review >](#)
- [13 Essential AWS IAM Best Practices >](#)
- [CIEM vs. IAM >](#)
- [Identity Security \[Cloud Edition\] >](#)
- [What is the Principle of Least Privilege \(PoLP\)? Use Cases, Benefits, and Implementation >](#)

Data Security in the Cloud

- [What is Cloud Data Security? Risks and Best Practices >](#)
- [What is a Data Risk Assessment? >](#)
- [What is Data Security Posture Management \(DSPM\)? >](#)
- [13 Essential Data Security Best Practices in the Cloud >](#)
- [Sensitive Data Discovery >](#)
- [Data Leakage: Risks, Causes, & Prevention >](#)
- [What are Data Security Controls? >](#)

Cloud Service Provider Security

- [Top 9 AWS Security Tools in 2025: Features and Limitations >](#)
- [AWS Security Best Practices: 10 Steps and How to Assess AWS Health >](#)
- [9 Best Practices for Stronger Azure Security >](#)
- [Native Azure Security Tools >](#)
- [Azure Security Risks & Mitigation Steps >](#)
- [AWS Security Risks >](#)
- [Google Cloud Security Best Practices >](#)
- [Top Google Cloud Security Tools \[By Use Case\] >](#)

Compliance

- [What is NIST Compliance? >](#)
- [What is Cloud Compliance? >](#)
- [A Comprehensive Guide to Navigating FISMA Compliance >](#)
- [HIPAA Cloud Compliance Essentials for Healthcare Providers >](#)
- [What is Cybersecurity Maturity Model Certification \(CMMC\)? >](#)
- [What Is Cloud Governance? Best Practices for A Strong Framework >](#)
- [What are CIS benchmarks? >](#)
- [ISO 27001 Controls: Fast Track Guide >](#)



PLATFORM

[Wiz CNAPP](#)
[Wiz Code](#)
[Wiz Cloud](#)
[Wiz Defend](#)
[Integrations](#)
[Environments](#)
[Documentation](#)

LEARN

[Customer stories](#)
[Train Your Team For Cloud](#)
[Blog](#)
[CloudSec Academy](#)
[Resources Center](#)
[Cloud threat landscape](#)
[Cloud Security Assessment](#)
[Vulnerability Database](#)

COMPANY

[About Wiz](#)
[Join the team](#)
[Newsroom](#)
[Events](#)
[Contact us](#)
[Trust Center](#)
[Our partners](#)

🌐 English (US) ▾



Assessment >

- What is CNAPP? [Cloud-Native Application Protection Platform] >
- What is CSPM? >

- Vulnerability Management Best Practices >
- Vulnerability Prioritization in the Cloud: Strategies + Steps >
- Risk-Based Vulnerability Management >
- Top OSS vulnerability management tools >

- 13 Essential AWS IAM Best Practices >
- CIEM vs. IAM >
- Identity Security [Cloud Edition] >
- What is the Principle of Least Privilege (PoLP)? Use Cases, Benefits, and Implementation >

Data Security in the Cloud

- What is Cloud Data Security? Risks and Best Practices >
- What is a Data Risk Assessment? >
- What is Data Security Posture Management (DSPM)? >
- 13 Essential Data Security Best Practices in the Cloud >
- Sensitive Data Discovery >
- Data Leakage: Risks, Causes, & Prevention >
- What are Data Security Controls? >

Cloud Service Provider Security

- Top 9 AWS Security Tools in 2025: Features and Limitations >
- AWS Security Best Practices: 10 Steps and How to Assess AWS Health >
- 9 Best Practices for Stronger Azure Security >
- Native Azure Security Tools >
- Azure Security Risks & Mitigation Steps >
- AWS Security Risks >
- Google Cloud Security Best Practices >
- Top Google Cloud Security Tools [By Use Case] >

Compliance

- What is NIST Compliance? >
- What is Cloud Compliance? >
- A Comprehensive Guide to Navigating FISMA Compliance >
- HIPAA Cloud Compliance Essentials for Healthcare Providers >
- What is Cybersecurity Maturity Model Certification (CMMC)? >
- What Is Cloud Governance? Best Practices for A Strong Framework >
- What are CIS benchmarks? >
- ISO 27001 Controls: Fast Track Guide >



PLATFORM

Wiz CNAPP
Wiz Code
Wiz Cloud
Wiz Defend
Integrations
Environments
Documentation

LEARN

Customer stories
Train Your Team For Cloud
Blog
CloudSec Academy
Resources Center
Cloud threat landscape
Cloud Security Assessment
Vulnerability Database

COMPANY

About Wiz
Join the team
Newsroom
Events
Contact us
Trust Center
Our partners

English (US) ▾

