perforce
Delphix

REPORT

The 2025 State of Data Compliance and Security Report

Contents

Executive Summary

The 2025 State of Data Compliance and Security Report offers critical insights into emerging trends, challenges, and priorities that large enterprises face in managing sensitive data. This year's research highlights how non-production environments pose increasing risks to compliance and security, and our team explores strategies organizations are deploying to mitigate these risks while fostering innovation.

Key Findings

1. Data breaches in non-production are on the rise.
   - **60% of organizations have experienced data breaches or theft** in non-production environments, an 11% increase from last year.
2. Static data masking is a leading compliance solution.
   - 95% of organizations are now leveraging static data masking.
3. Organizations are investing in AI data privacy.
   - **86% of organizations plan to invest in AI data privacy over the next 1-2 years.**
4. There's mass confusion about the exposure of sensitive data in AI model training.
   - **91% say sensitive data should be allowed in AI training and testing.**
   - 82% believe it is safe to do so.
   - **Yet, 78% are highly concerned about theft or breach of model training data.**
   - And 68% worry about privacy and compliance audits.
5. Compliance exceptions create security and compliance gaps.
   - **84% of organizations allow compliance exceptions in non-production environments.**
6. All organizations (100%) have data in non-production that is subject to privacy regulations.

Ann Rosen
Director of Product Marketing

Steve Karam
Principal Product Manager

Ross Millenacker,
Senior Product Manager

A Letter From the Authors

Data compliance and security are critical for enterprise success, especially in non-production environments, which are often overlooked despite their heightened exposure risks. As organizations face evolving data privacy regulations and growing security threats, the stakes have never been higher.

Last year, we launched the State of Data Compliance and Security Report to shed light on the challenges posed by sensitive data in non-production. This year's report highlights that the growth of such data shows no signs of slowing, with increasingly severe consequences for organizations.

Non-production environments — which mirror production settings and support analytics, AI development, and testing — remain vulnerable. Manual processes and fragmented tools lead to delays, inefficiencies, and weakened security. **Alarmingly, 60% of organizations reported breaches or theft in non-production environments in the past year.**

Additionally, the vast majority of organizations confirmed increased data exposure due to expanding AI and analytics initiatives. The need for robust, automated solutions to manage data privacy is clear. Positively, most organizations plan to invest in AI data privacy in the next 1-2 years, signaling a shift toward proactive measures.

The 2025 State of Data Compliance and Security Report, based on insights from 280 enterprise leaders, explores key trends, risks, and solutions for securing sensitive data in non-production. It emphasizes strategies like automated, irreversible static data masking to reduce risks, ensure compliance, and foster innovation.

We hope this report helps your organization enhance data compliance and security. Together, we can build a safer, more secure future.

Sincerely,

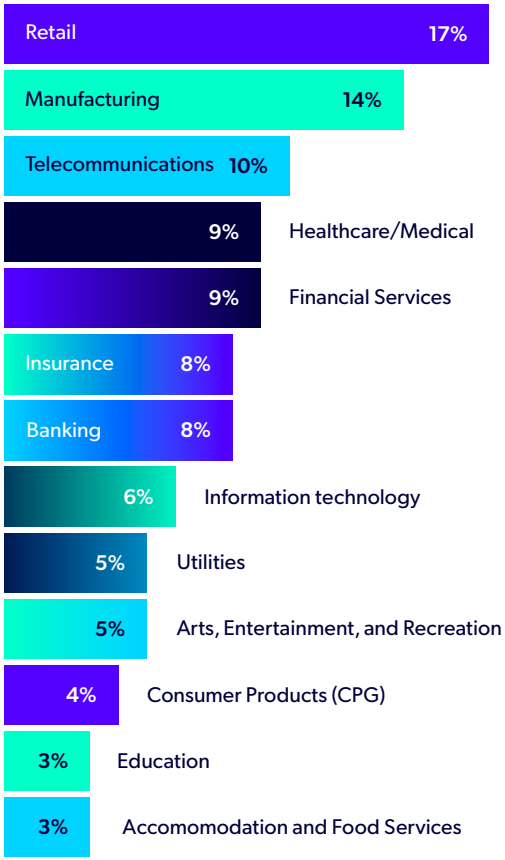
Ann Rosen, Steve Karam, Ross Millenacker

Who Took the Survey?

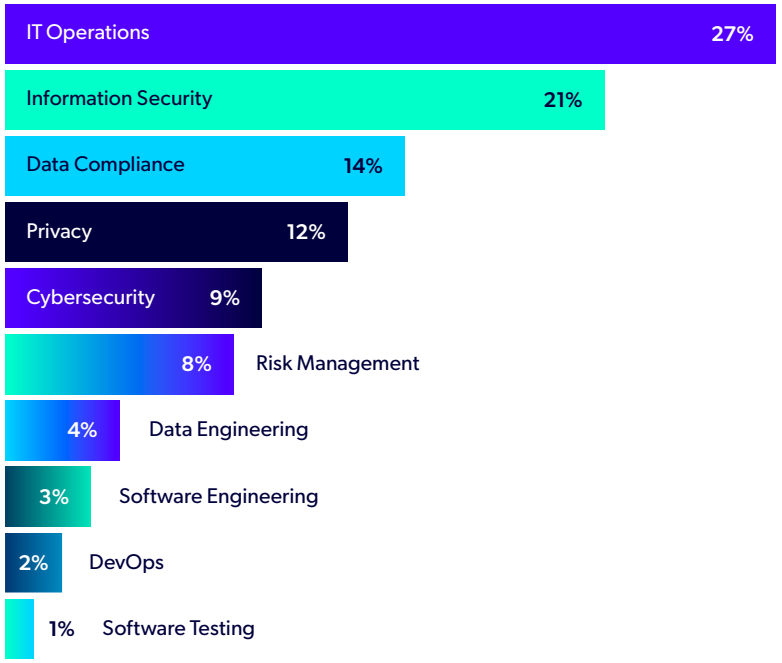
Perforce Delphix partnered with a third-party research firm again this year to run a survey that was sent to **280 enterprise leaders** around the globe. This year, all respondents were in leadership positions, and they spanned a breadth of job functions — primarily IT, InfoSec, and data compliance or privacy. Many were in industries such as financial services, retail, manufacturing, telecoms, healthcare and insurance, and more.

Find full demographics on pages 33-34.

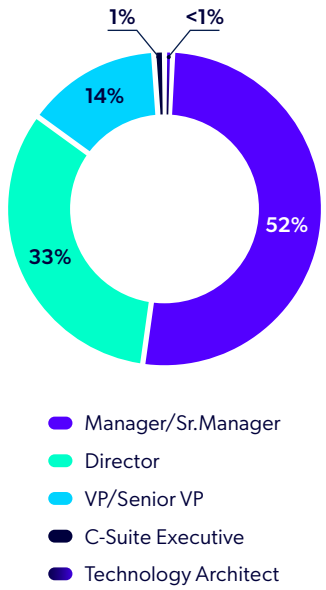
INDUSTRY



JOB FUNCTION



JOB ROLE



Sensitive Data Exposure Continues to Proliferate Across Non-Production Environments

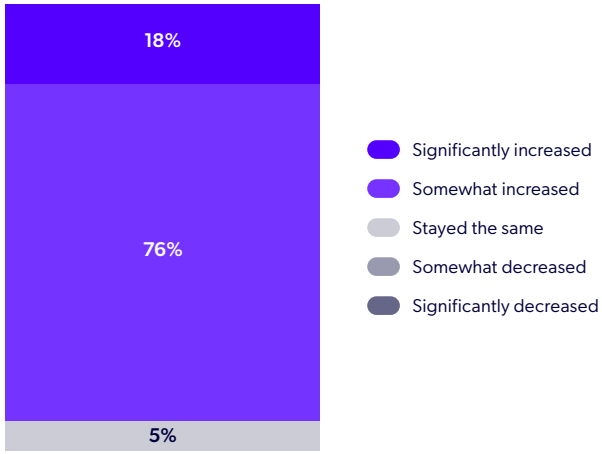
We surveyed people whose organizations work with sensitive consumer data, including personally identifiable information (PII), protected health information (PHI), and financial details.

What we found: Sensitive data within non-production environments has reached unprecedented levels.

According to the 2025 survey, an overwhelming **95% of organizations report storing more sensitive data in non-production environments** compared to the previous year. This is a 27% increase from last year's results, when 75% of respondents said that volume had increased year-over-year. In fact, 18% said it significantly increased.

Only 5% said it had stayed the same, and no one reported a decrease.

HOW HAS THE OVERALL VOLUME OF SENSITIVE DATA THAT YOUR ORGANIZATION STORES IN NON-PRODUCTION ENVIRONMENTS CHANGED OVER THE PAST 12 MONTHS? (NUMBER OF RECORDS)

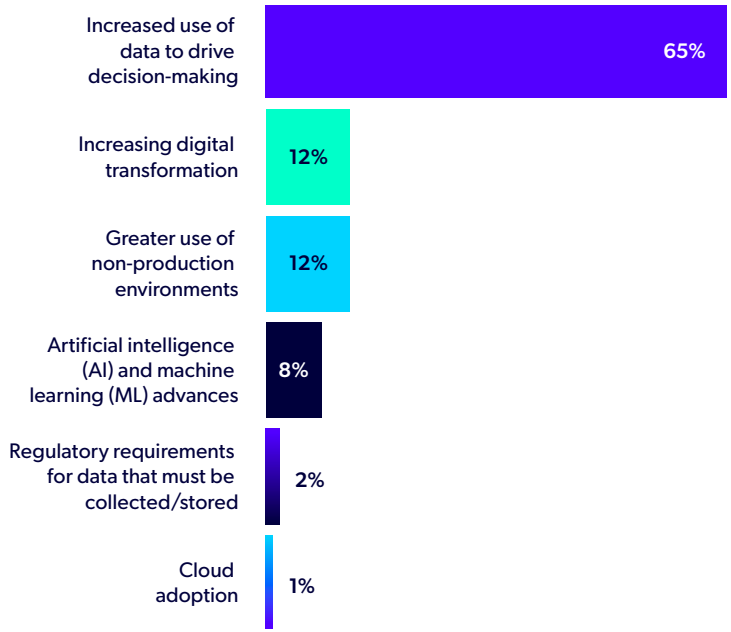


As sensitive data spreads across non-production, the risks around that data increase, too — as evidenced by the alarming number of organizations who have experienced adverse effects. We will explore this more in a later section. (See page 10.)

This significant increase is cause for concern. But why is it happening?

By far, **the biggest reason is the increased use of data-driven decision-making, according to 65% of respondents.** That's 33% higher than last year, when just 49% attributed it to this reason.

WHAT DO YOU BELIEVE IS THE BIGGEST REASON THAT THE OVERALL VOLUME OF SENSITIVE DATA THAT YOUR ORGANIZATION STORES IN NON-PRODUCTION ENVIRONMENTS HAS INCREASED? (NUMBER OF RECORDS)



Enterprises need data to drive decision-making, and they need it now, leading to the proliferation of data across the organization.

Sometimes the consumers of PII/PHI analytics reports and dashboards do need to view the actual, unchanged sensitive data. But for many analytics teams, there is no need to view the sensitive data itself; it should be masked.

Leaders can't sacrifice the security and protection of that data for speed. They need both, without any trade-offs.

Where is Sensitive Data Being Used?

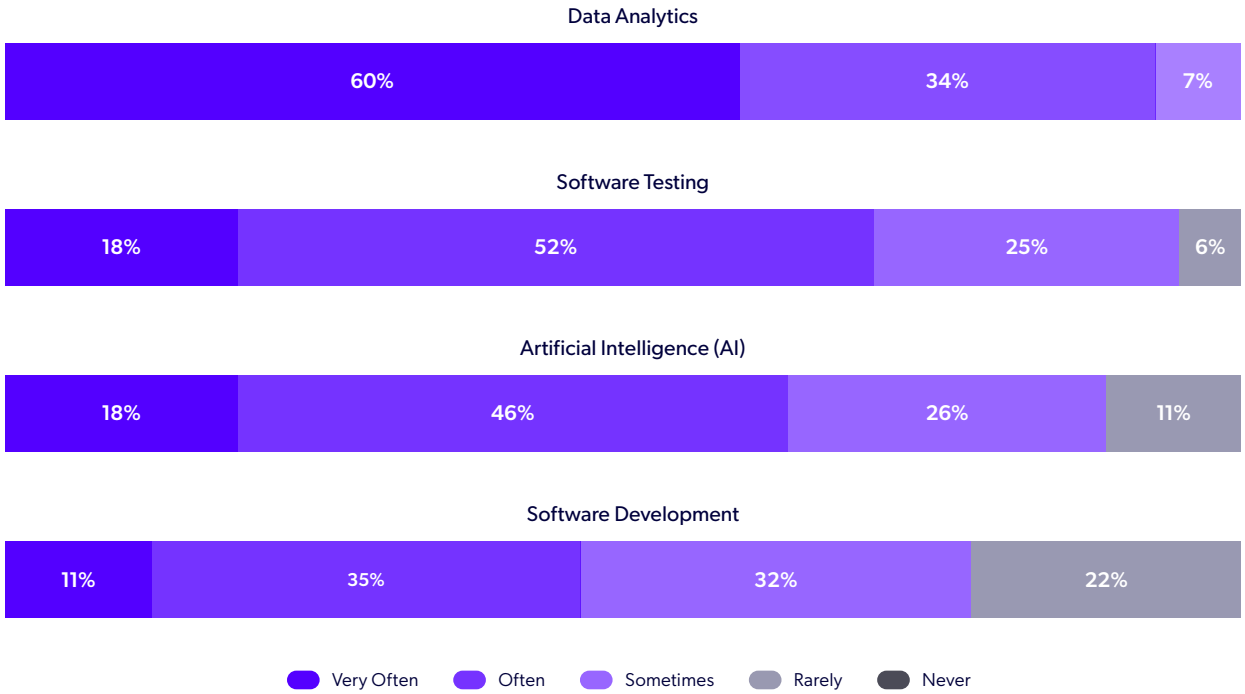
We asked respondents how often their organization works with sensitive data in various kinds of environments.

In line with the finding about an increase in data-driven decision-making, the most common environment was data analytics, followed by software testing and artificial intelligence. Software development was the last category, with the largest percentage of respondents reporting only "sometimes" using sensitive data in it.

No organizations reported that they "never" use sensitive data in data analytics, software testing, AI, or software development.

HOW OFTEN DOES YOUR ORGANIZATION WORK WITH SENSITIVE DATA IN THE FOLLOWING ENVIRONMENTS?

(E.G., PII/PHI SUCH AS SOCIAL SECURITY NUMBERS, ADDRESS, IDENTIFYING HEALTH INFORMATION, PERSONAL FINANCIAL INFORMATION)



100% Use Sensitive Data in Analytics

The most common environment by far was data analytics, which 100% use sensitive data in. Last year, this number was 99%. Additionally, 93% say they use it "often" or "very often."

This finding reflects the critical role data analytics plays in shaping data-driven decision-making. But securing sensitive data in analytics workflows is crucial for maintaining compliance, building trust, and ensuring organizations can safely leverage valuable insights from their data.

95% Use Sensitive Data in Software Testing

The second most common environment was software testing, which 95% use sensitive data in. That's down a little bit from 97% last year. 70% use it "often" or "very often."

Software testing is a critical weak point for many organizations, with rapid development cycles increasing the risk of human error and vulnerabilities. To reduce risks, it's crucial to ensure that test environments use compliant, secure data. But for testing to be reliable and meaningful, it's also critical for that data to be consistent across tables, sources, and environments.

90% Use Sensitive Data in Artificial Intelligence

The third most common environment was artificial intelligence, which 90% use sensitive data in. Last year, 93% reported working with sensitive data in AI. 64% use it "often" or "very often" in these environments.

This is an alarmingly high number. It is extremely risky to use sensitive data in these environments due to the likelihood of it reappearing in unexpected places. AI models don't forget inputs, so sensitive customer information should never enter these pipelines in the first place. We will explore the implications of using sensitive data in AI environments in more detail later on.

78% Use Sensitive Data in Software Development

This year, 78% of respondents said they work with sensitive data in software development, with 46% using it "often" or "very often." (Last year, these numbers were 95% and 81%, respectively.)

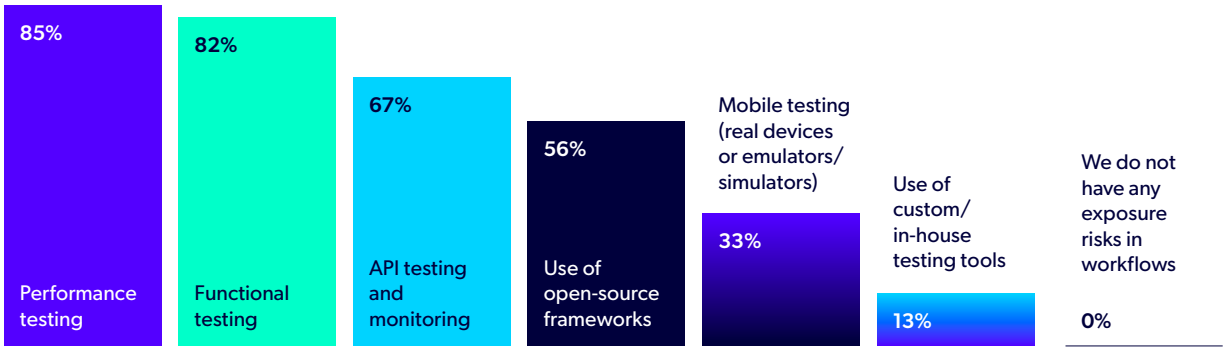
These numbers remain quite high. No teams should be using sensitive data in development environments because of the security and compliance risks this practice poses.

We asked respondents where in their software testing workflows they believe their organization is most at risk of exposing sensitive data in non-production environments. **All (100% of) respondents reported some level of exposure,** with no organizations indicating zero risk. The highest exposure points were performance testing (85%), functional testing (82%), and API testing and monitoring (67%), indicating that core quality assurance activities remain a consistent source of potential data leakage.

These results suggest that even mature testing practices struggle to fully eliminate sensitive data from high-value workflows. Enterprise teams can reduce exposure by masking sensitive data before it enters testing environments, as well as by applying environment isolation and strengthening API monitoring protocols. Such measures allow organizations to maintain test coverage and speed while limiting security and compliance risks.

100% Have Exposure Risk in Software Testing

WHERE IN YOUR SOFTWARE TESTING WORKFLOWS DO YOU BELIEVE YOUR ORGANIZATION IS MOST AT RISK OF EXPOSING SENSITIVE DATA IN NON-PRODUCTION ENVIRONMENTS? SELECT ALL THAT APPLY.

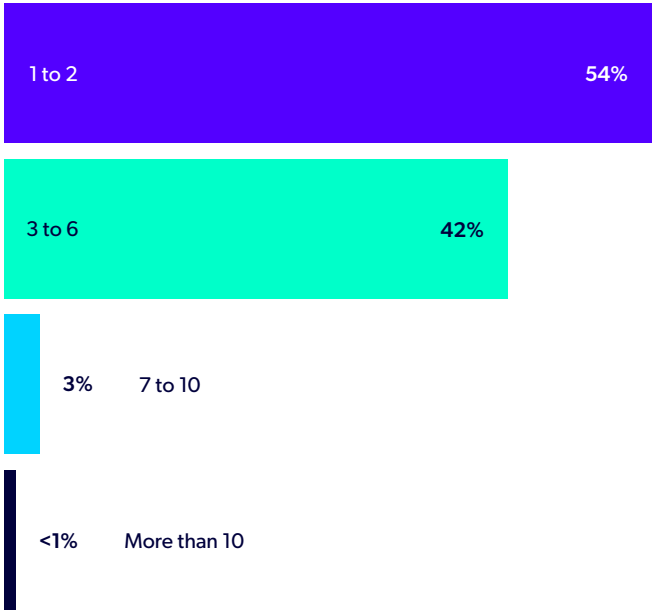


Sensitive Data Spreading to Non-Production Magnifies the Risk Footprint

Organizations frequently maintain multiple copies of sensitive data to support use cases like software development, testing, and analytics. **Almost half of respondents (45%) reported that for each dataset they have in production, they have 3-10 copies of that dataset in non-production.**

FOR EACH DATASET IN YOUR PRODUCTION APPLICATIONS, HOW MANY COPIES DO YOU TYPICALLY MAINTAIN IN NON-PRODUCTION?

(E.G., DEVELOPMENT, TESTING, ANALYTICS, OR AI)



We found that more than half (54%) report having just 1-2 copies of a dataset in non-production for every one in production. But in our experience, this is not by choice. It's often resource constraints that limit businesses to fewer environments.

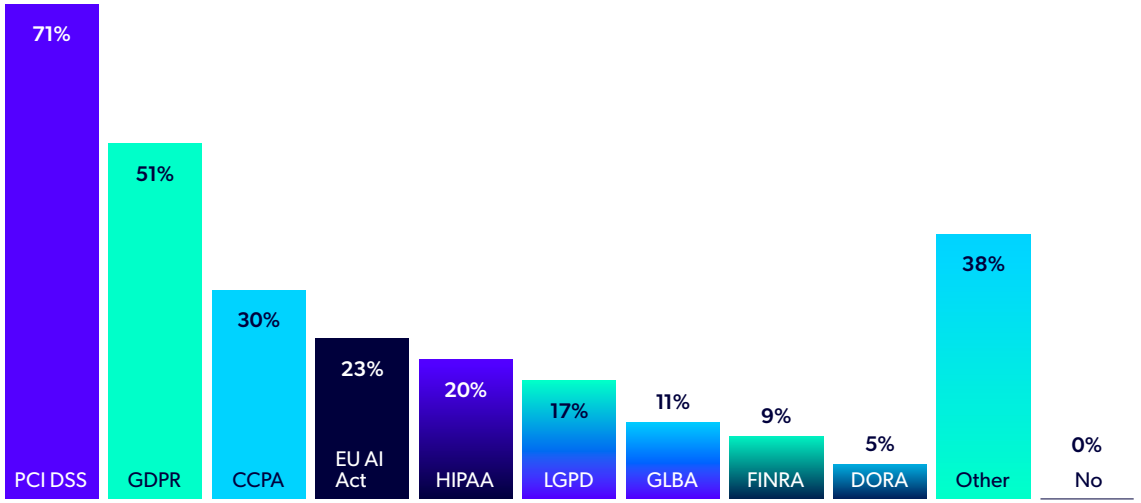
When it's faster and easier to spin up additional test environments, many organizations who start with 1-2 would prefer to expand to 3-6, 7-10, or even more. In fact, recently, we at Perforce Delphix helped one company with dozens of development teams spin up hundreds of ephemeral test environments, each one available on-demand and purpose-built for each test.

By using highly cost-effective data virtualization approaches, organizations can efficiently and quickly spin up many copies of their data at almost no additional cost.

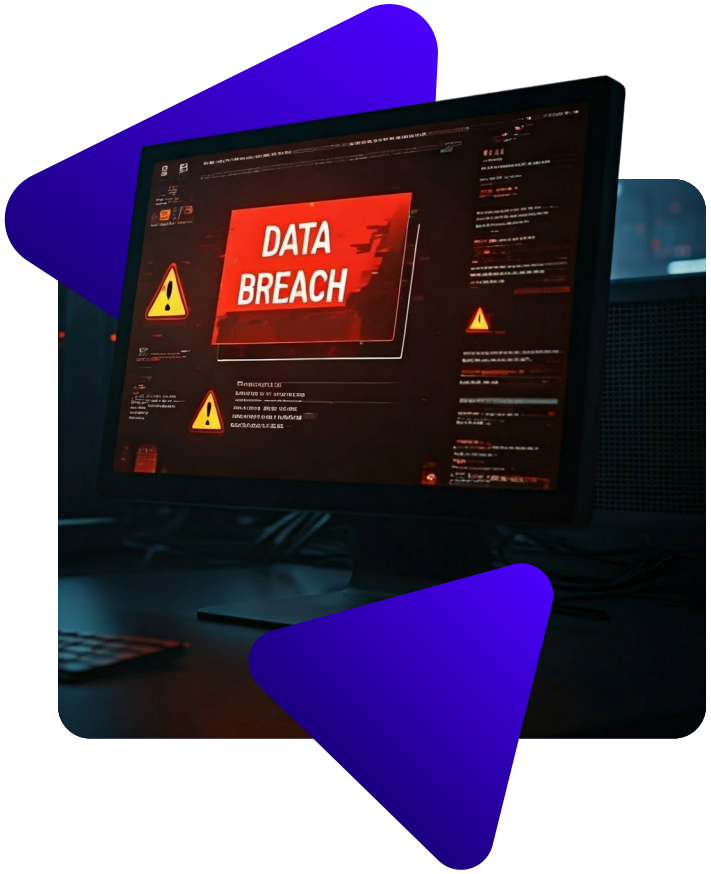
100% Have Data Subject to Privacy Regulations in Non-Production

All surveyed organizations reported having data that is subject to privacy regulations in their non-production environments. The top regulations cited were PCI DSS (71%), followed by GDPR (51%) and CCPA (30%).

DOES YOUR ORGANIZATION HAVE DATA IN NON-PRODUCTION ENVIRONMENTS THAT IS SUBJECT TO ANY OF THE FOLLOWING DATA PRIVACY REGULATIONS?



This underscores the elevated risk profile associated with non-production environments. Because non-production typically does not have the same controls and governance as production, extra attention should be given to the data in these environments. For example, sensitive data should always be masked in non-production to render it useless to cybercriminals — and to ensure compliance.

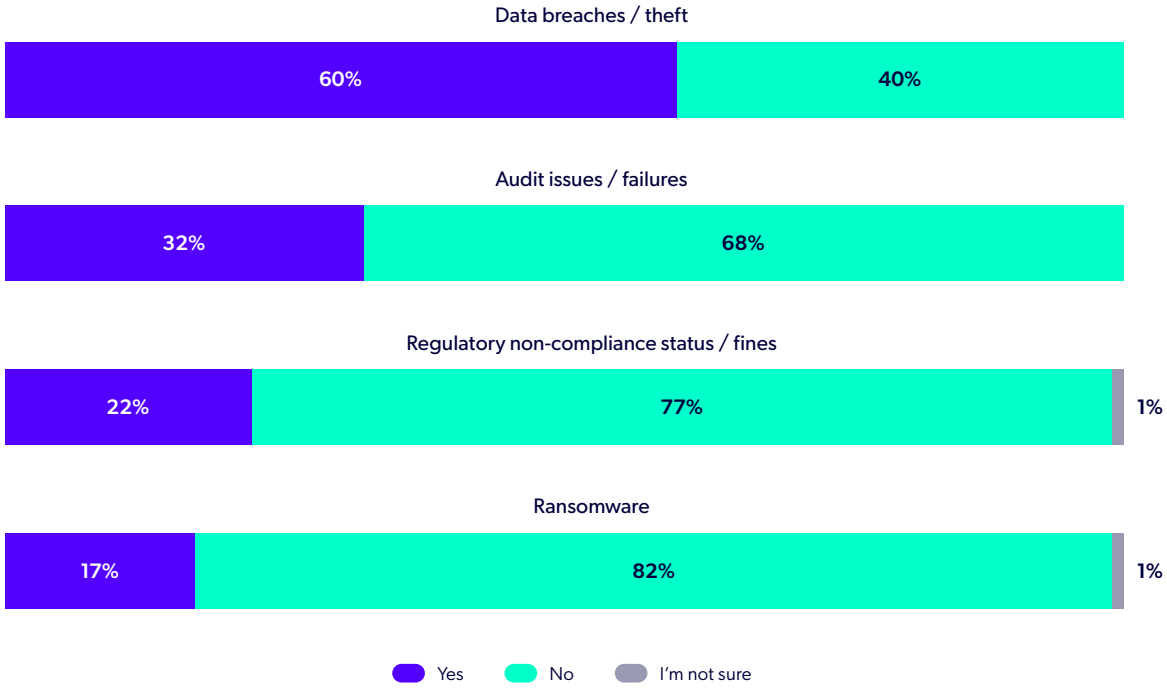


Most Organizations Have Suffered an Attack on Non-Production Data. Fear is at an All-Time High.

Here's an alarming fact:

60% of organizations have already experienced breaches or theft in non-production environments. That's **up 11%** from **54% last year.**

HAS YOUR ORGANIZATION EXPERIENCED ANY OF THE FOLLOWING INVOLVING SENSITIVE DATA IN NON-PRODUCTION ENVIRONMENTS?



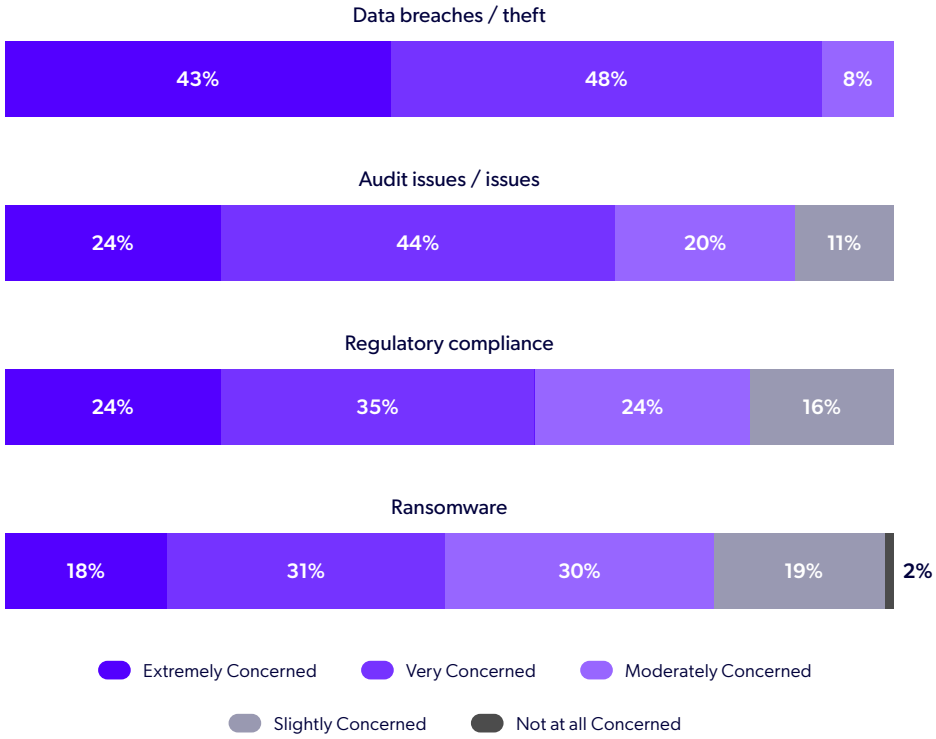
"The fact that 60% of organizations experienced data breaches or theft in non-production environments is a wake-up call. In highly regulated industries, the same rigor we apply to securing production systems must extend to dev and test environments — especially as they often house sensitive data and critical integrations. As we accelerate cloud adoption and AI development, securing non-prod environments isn't just a best practice — it's a business imperative."

— Arvind Anandam,
Sr. Dir, Cloud and Platform Engineering,
Worldpay

worldpay

It's no wonder a **whopping 99% of respondents are at least "moderately" concerned about data breaches and theft in non-production** — with close to half of them (43%) being "extremely concerned." (Last year, just 30% reported being extremely concerned.)

WHAT IS YOUR LEVEL OF CONCERN WITH THE FOLLOWING INVOLVING SENSITIVE DATA IN NON-PRODUCTION ENVIRONMENTS?



There is deepening apprehension among security leaders about this overlooked yet critical area of their data estates. And their concern is valid. Protecting environments where data is constantly growing is a strain on resources.

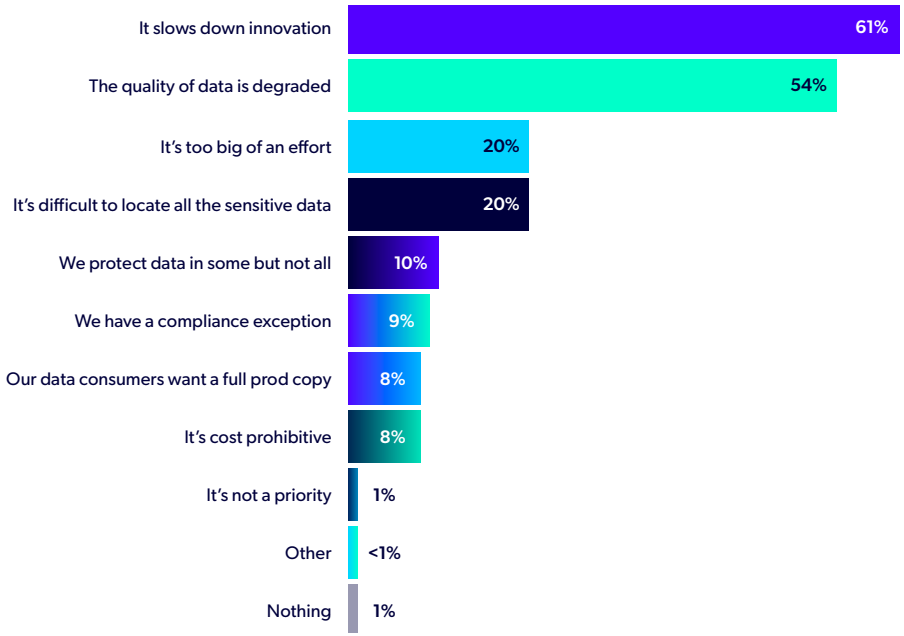
Thankfully, there are approaches you can take to alleviate these concerns and reduce the strain — see page 31 to learn more.

Speed and Quality are the Biggest Barriers to Protecting Data in Non-Production

With so much at stake, what is preventing organizations from protecting all sensitive data in non-production environments?

Overwhelmingly, leaders told us: because it slows down innovation and it lowers data quality.

IDEALLY ORGANIZATIONS WANT TO IRREVERSIBLY PROTECT ALL SENSITIVE DATA IN NON-PRODUCTION ENVIRONMENTS. WHAT IS PREVENTING YOUR ORGANIZATION FROM PROTECTING ALL YOUR SENSITIVE DATA IN NON-PRODUCTION?



61% Fear Innovation Slowdowns

Protecting sensitive data is often seen as a hindrance to innovation. For many organizations, it can take weeks to provision production databases to non-production environments. Then, it can be difficult to locate all sensitive data (as 20% of respondents reported), since it can appear in various formats from various sources.

Masking data is incredibly time-consuming when using manual and other sub-optimal methods, adding weeks to the timeline. The problem is further exacerbated when dealing with data sources that are larger than 10TB. And finally, delivering masked data to downstream teams can itself be a bottleneck, especially if multiple copies of the dataset are needed.

All of these hurdles may be why 20% said protecting sensitive data is "too big of an effort."

With all of that in mind, it's no wonder 61% of respondents said protecting data is a barrier to innovation — especially considering how many organizations are ramping up AI, as well as investing in AI and data privacy. (More on this later. See page 25.) But, as we will explore later, data compliance doesn't have to bottleneck innovation.

54% Fear Lower Data Quality

Almost all leaders believe that protecting sensitive data in non-production involves trade-offs. (We'll touch on this in a bit.) **This year, 54% of respondents cited "the quality of data is degraded" as a barrier to protecting sensitive data** — a growing concern for organizations aiming to maintain both security and software quality.

Because respondents could select multiple answers, it's possible that their concern about data quality being too low goes hand-in-hand with their concern about hindering innovation. High-quality data allows for more accurate and actionable BI and faster, higher quality software releases.

Protecting data is a necessary safeguard, but it doesn't come without challenges. Altering fields to remove sensitive data, without considerations for data realism, can lead to failed testing. Maintaining **referential integrity** across interdependent datasets makes the process even more complicated. For example, if a customer name is inconsistently masked across systems, integration testing breaks down. And if development and QA teams test with low-quality datasets, it increases the likelihood of defects downstream.

To minimize risks and maintain high software quality, organizations must prioritize consistent and realistic data masking throughout their pipelines. It is possible to deliver quality, compliant data to downstream teams when and where they need it to shift left and minimize software defects.

NOTE

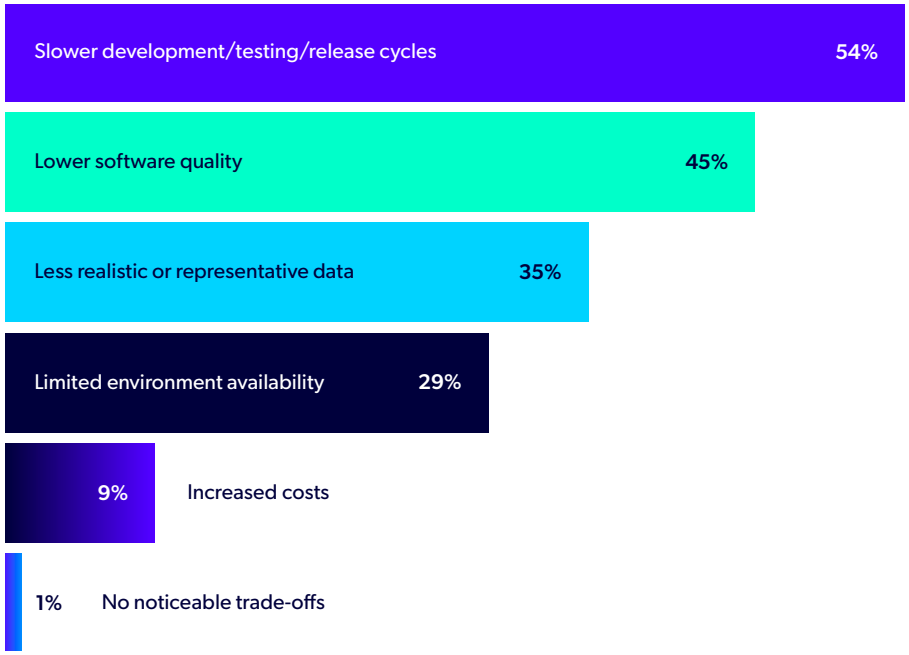
It's important to maintain referential integrity across related database tables — meaning if a value appears in multiple places, the masked value should remain consistent.

The Perceived Trade-offs

Knowing leaders are highly concerned with speeding up innovation and maintaining data quality, it makes sense that they consider the top trade-offs when protecting sensitive data in non-production to be:

- **Slower development/testing/release cycles (54%)**
- **Lower software quality (45%)**
- **Less realistic or representative data (35%)**
- **Limited environment availability (29%)**

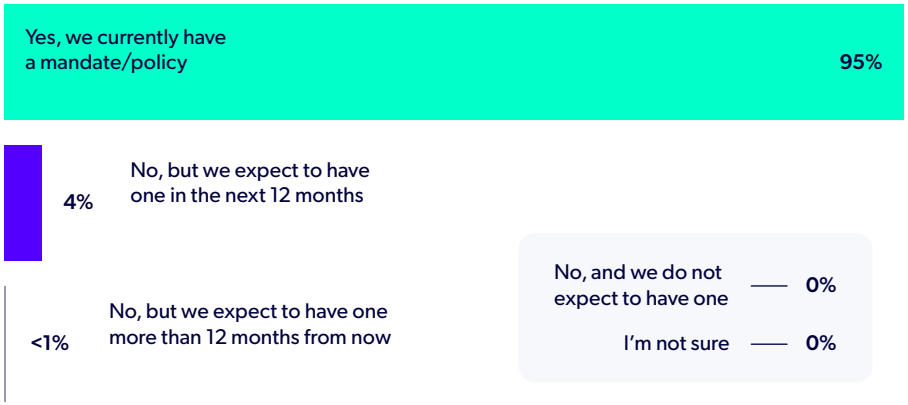
WHAT TRADE-OFFS HAS YOUR ORGANIZATION EXPERIENCED WHEN IMPLEMENTING SENSITIVE DATA PROTECTION IN NON-PRODUCTION ENVIRONMENTS?



Most Organizations Have Masking Mandates or Policies

The majority of organizations (95%) have a masking mandate or policy in place for non-production environments. This reinforces our findings that organizations already have high concern about sensitive data vulnerabilities in non-production. But what are they doing about it? We will explore this later.

DOES YOUR ORGANIZATION HAVE A DATA MASKING MANDATE/POLICY IN NON-PRODUCTION ENVIRONMENTS?

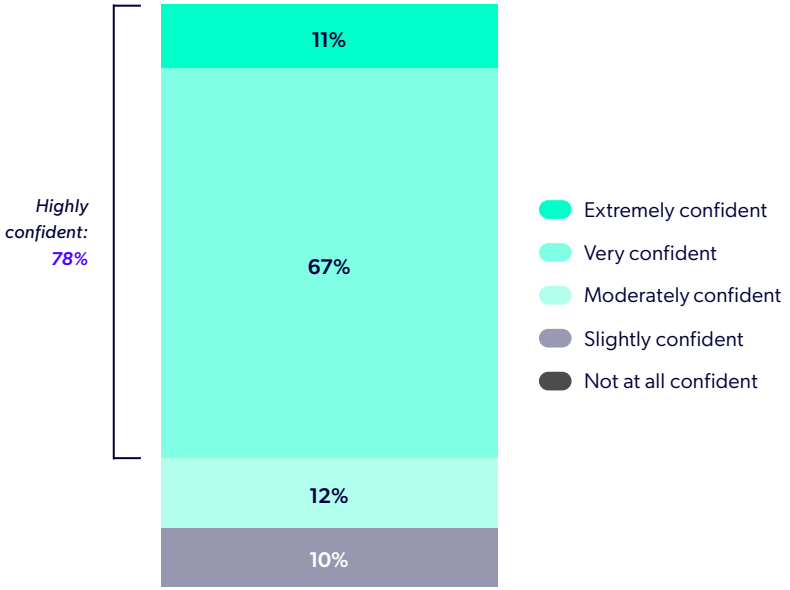


Most Organizations are Confident They Can Identify Sensitive Data

A high number of respondents — 90% — reported feeling confident in being able to identify all sensitive data in non-production environments. Of those, 78% are either "very" or "extremely" confident.

HOW CONFIDENT ARE YOU THAT YOUR ORGANIZATION IS ABLE TO IDENTIFY ALL SENSITIVE DATA RESIDING IN NON-PRODUCTION ENVIRONMENTS?

(E.G., DISCOVER, PROFILE)



We found this surprising, especially considering 100% of respondents have sensitive data that is subject to regulations in non-production — and considering the majority of organizations reported the volume of sensitive data in non-production has increased over the past 12 months.

Perhaps there is a semantic question here. Organizations may know in what systems they have sensitive data. But how easy is it to find all sensitive data interspersed throughout complex and heterogenous data systems? For example, say you have a point-of-sale application with a "customers" and "orders" pairing of databases. You know there will be PII interspersed across tables, but you don't know the specific fields, and you may not know where the PII exists in the database schema. Manually scanning for this data is highly error-prone.

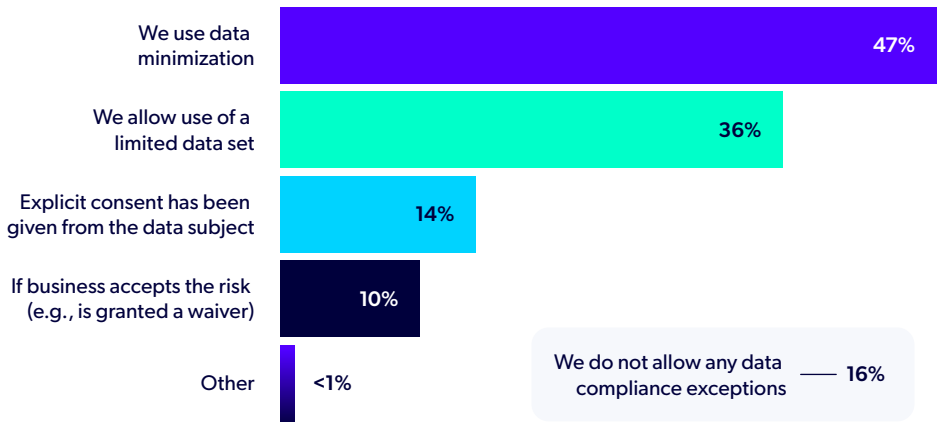
Our regular conversations with business leaders suggest that they do struggle with and are concerned about finding all sensitive data across non-production.

Yet, 84% of Organizations Allow Compliance Exceptions

Despite the fact most organizations have data masking policies or mandates and feel confident they can identify all sensitive data in non-production — and despite the high level of concern over exposing this data to breach, data theft, compliance audits, and more — 84% allow data compliance exceptions in non-production.

84% allow data compliance exceptions in non-production

DOES YOUR ORGANIZATION ALLOW ANY DATA COMPLIANCE EXCEPTIONS IN NON-PRODUCTION ENVIRONMENTS?



This shocked us. Knowing that 60% of organizations have experienced data breach or theft in non-production, it is alarming that so many organizations continue to allow compliance exceptions.

But we got a sense of why it's happening from the responses to our questions about protecting sensitive data in non-production environments. Protecting data is perceived to be difficult and disruptive, an incredible roadblock and hindrance to innovation.

Fortunately, it is possible to achieve speed, quality, and compliance — without trade-offs. We will explore this more in a later section, "Accelerate Data Compliance with Perforce Delphix."

95% of Organizations Use Static Masking

The rapid expansion of sensitive data across non-production environments introduces a host of vulnerabilities for organizations, and protecting this data can feel daunting.

To combat these challenges, the majority (95%) of organizations have turned to static data masking as a key solution. This method allows teams to safeguard sensitive information while maintaining compliance, effectively mitigating exposure in non-production environments.

Static masking not only addresses data security concerns but also aligns with industry regulations, making it an essential practice for reducing risk.

95% of organizations use static masking

WHAT DO YOU USE FOR PROTECTING SENSITIVE DATA IN NON-PRODUCTION ENVIRONMENTS?

(e.g., irreversibly replaces sensitive data values with fictitious, yet realistic equivalents that reflect production data and its complexity and relational attributes

Static masking's widespread use underscores its effectiveness in protecting sensitive data and supporting critical activities in non-production environments like software testing. By eliminating data vulnerabilities and ensuring compliance with privacy regulations, static data masking enables organizations to maintain data quality while driving operational efficiency and security in non-production settings.

"It's encouraging to see widespread adoption of static data masking, especially as more organizations recognize the risks of using sensitive data in non-production environments. But masking alone isn't enough — we need to combine it with strong access controls, auditability, and, where possible, synthetic data to truly minimize exposure and maintain compliance in healthcare IT operations."

CalOptima Health

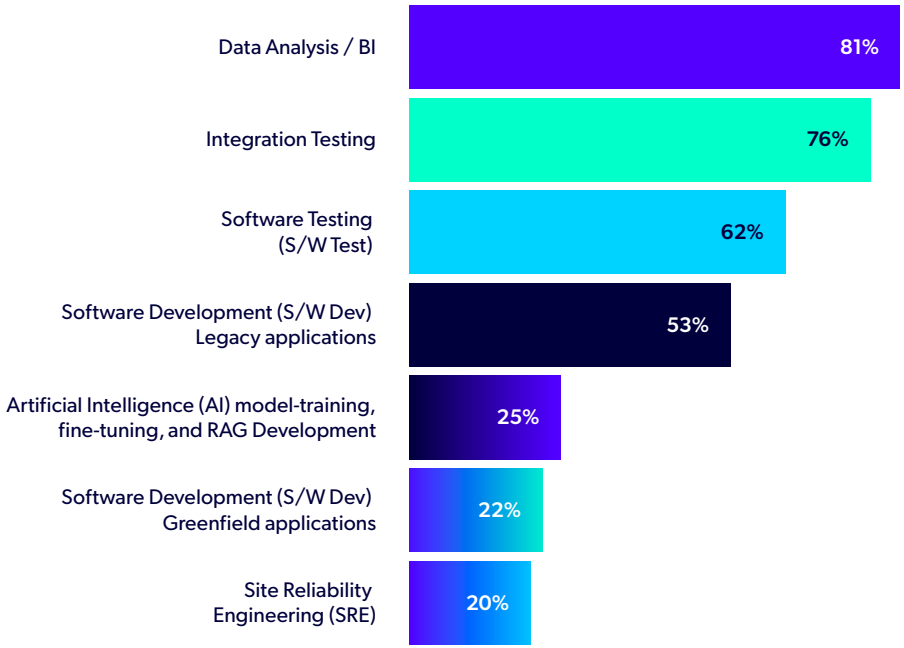
— Clare Xia, ITS Manager
CalOptima Health

How Static Masking is Used

We asked which non-production environments organizations are using static data masking in. The top responses were:

- Data analysis / BI (81%)
- Integration testing (76%)
- Software testing (62%)
- Software development / Legacy applications (53%)

IN WHICH NON-PRODUCTION ENVIRONMENTS IS YOUR ORGANIZATION USING STATIC DATA MASKING TO PROTECT SENSITIVE DATA? SELECT ALL THAT APPLY.



81% Use it For Data Analysis & BI

Sensitive data is at the core of business insights. Static masking lets analytics and intelligence teams strike a balance between usability and security. It protects sensitive data in the development of analytics workflows, such as offshore development of reports and dashboards or ETL workflows.

NOTE

Static data masking is the best way to ensure referential integrity across tables, schemas, databases, and cloud environments.

76% Use It For Integration Testing

Integration testing requires data from various interconnected systems. When masking data for integration testing, the referential integrity of data is critical. Static masking must ensure that sensitive data is consistently masked. This lets teams validate system interactions securely without exposing sensitive information.

62% Use It For Software Testing

Software testing and development requires realistic, production-like datasets to identify bugs and optimize performance. Just like with integration testing, consistent data is key, since it ensures tests are repeatable and reliable. Static masking provides secure versions of these datasets, reducing risk while allowing teams to perform evaluations in a controlled, production-like, compliant environment.

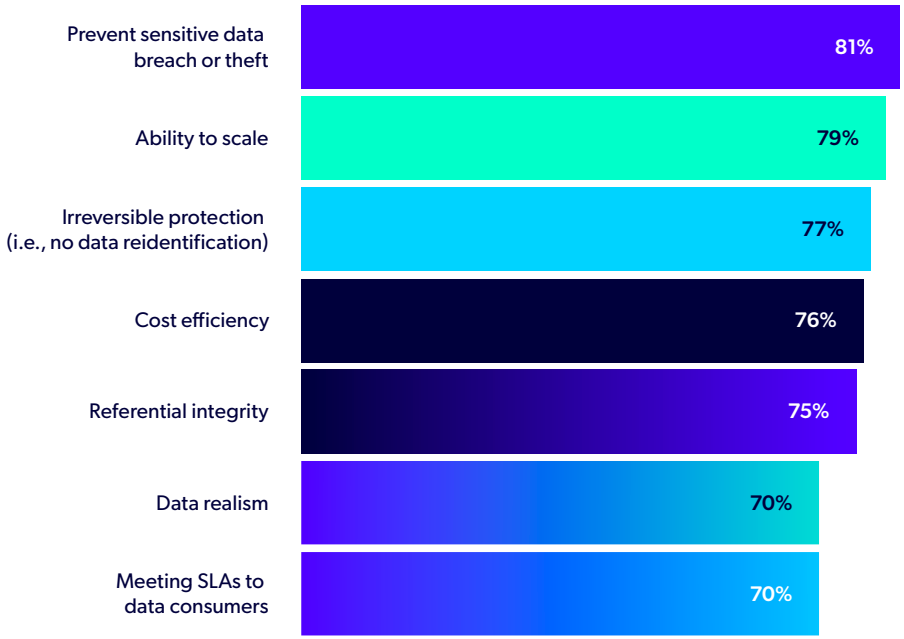
Key Qualities of Static Masking

We also asked respondents how effective they think static masking is across various dimensions like scale, cost, speed, and referential integrity. Most rated it as "very" or "extremely" effective in these key dimensions:

- Preventing sensitive data breach or theft (81%)
- Scalability (79%)
- Cost efficiency (76%)

HOW EFFECTIVE IS STATIC DATA MASKING AT THE FOLLOWING FOR PROTECTING SENSITIVE DATA IN NON-PRODUCTION ENVIRONMENTS?

Rated 'Very' or 'Extremely' Effective



81% Cite Effectiveness for Preventing Data Breach or Theft

Unlike reversible masking types (like dynamic masking), static masking replaces sensitive values with fictitious yet realistic equivalents so it can't be exposed in downstream environments. It doesn't surprise us that preventing data breach or theft is seen as a standout quality.

79% Cite Effectiveness for Scale

When implemented correctly, static masking can scale across large, complex enterprise data environments. High-performance techniques like parallel processing and automation, combined with efficient delivery methods such as data virtualization, can drastically reduce the time to deliver compliant data — from months to hours or even minutes. Once masked, the data can be reused across multiple environments, including development, testing, and training.

76% Cite Effectiveness for Cost Efficiency

Static masking's scalability also makes it highly cost effective. It eliminates the need for repeated, manual, ad hoc masking processes. Some best practices for cost efficient masking workflows include:

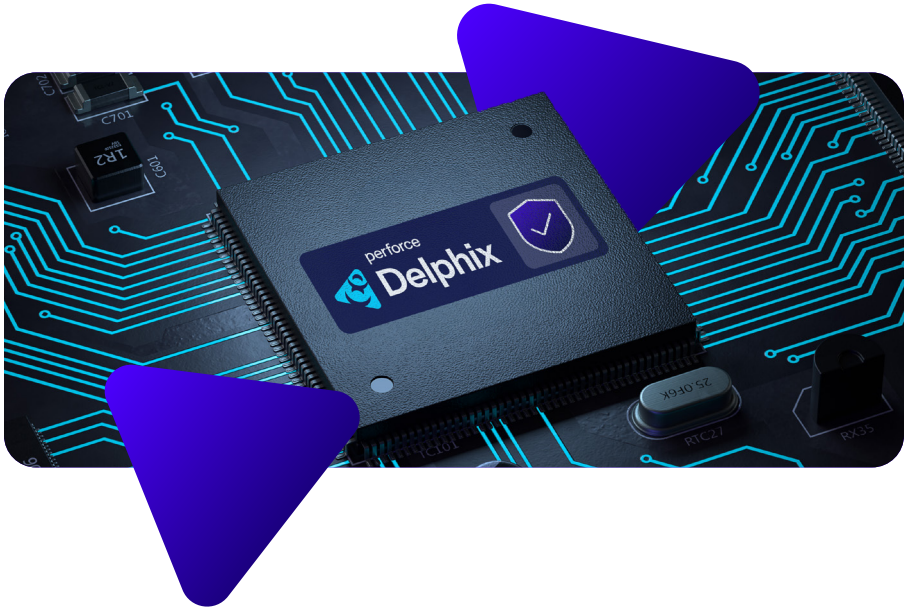
1. Create secure, compliant copies from a statically-masked gold copy of a database.
2. Automatically apply masking policies across datasets.
3. Use ephemeral data for testing.
4. Implement cost-effective delivery, like virtualization.
5. Integrate with existing workflows, e.g., DevOps toolchains.

Transforming Data Privacy Compliance with Delphix Static Data Masking

Organizations handling sensitive data face constant challenges to ensure privacy, compliance, and security. Delphix static data masking offers a scalable solution for high-quality, compliant data. **With Delphix, 77.2%* more data and data environments are masked and protected.**

Delphix provides production-like masked data that maintains referential integrity, enabling secure, high-quality data for development and testing, analytics, and AI. Unlike with other masking types, data that is masked with Delphix is irreversible. It protects against breaches and re-identification risks, helping enterprises meet compliance standards while driving innovation.

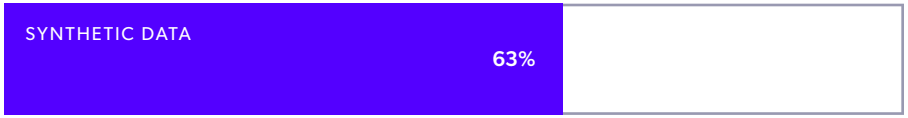
Learn more about Delphix on page 31.



63% of Organizations are Using Synthetic Data

We also found that 63% of organizations are now using synthetic data in some capacity. We're not surprised by this finding.

WHAT DO YOU USE FOR PROTECTING SENSITIVE DATA IN NON-PRODUCTION ENVIRONMENTS?



(e.g., synthetically generated,net-new data that may or may not provide data realism and referential integrity vis-a-vis production data)

As a provider of both synthetic data and static data masking solutions, Perforce has a front row seat to observe dynamics in the market for data compliance solutions. Over the past year we have noticed a subtle change — more organizations are exploring synthetic data solutions.

It's also worth noting a significant overlap; Many people are using synthetic data alongside static data masking. More on this later.

We suspect that synthetic data mandates, as opposed to the broader data compliance mandate mentioned above, may play a role. The rise of AI is also giving rise to the use of synthetic data. If and when organizations aim to protect data in AI workflows, synthetic data is emerging as a common approach. (Spoiler alert: We uncovered mass confusion related to protecting data in AI, which we will explore later in this report.)

A Note on Synthetic Data

Before we dive into more findings around synthetic data, we should discuss **what we mean when we talk about synthetic data.**

To some, synthetic data means a script for generating a small and basic dataset to test a new piece of code, or even asking ChatGPT to generate test data. These are bare-bones approaches, but do they address enterprise-level compliance challenges? Likely not.

There are "AI-native" synthetic data solutions that are based on AI-generated data. In our experience, these are primarily used in AI development. However, questions remain about their scalability and realism compared to production data.

Most interestingly, other synthetic data solutions are effectively static data masking in disguise. These solutions replace real data with fake data, whereby all data is replaced instead of just the sensitive identifying columns.

As you read the following findings, keep in mind that synthetic data means different things to different people.

How Is Synthetic Data Used

We asked the 63% of organizations who use synthetic data in which non-production environments they are using it.

79% Use Synthetic Data in Analytics

IN WHICH NON-PRODUCTION ENVIRONMENTS IS YOUR ORGANIZATION USING SYNTHETIC DATA?

We expected synthetic data to be most popular in AI environments. We learned that the top use case for synthetic data is actually analytics (79%). While we do not expect synthetic data to be used in production reports and dashboards, the reason this response is so high may be because synthetic data is used for the development of reports and dashboards or ETL workflows, all of which are often undertaken by offshore development teams.

74%+ Use Synthetic Data for Software Testing and Development

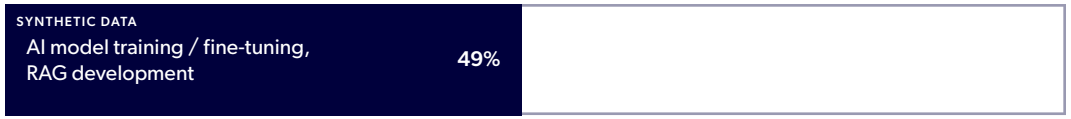
IN WHICH NON-PRODUCTION ENVIRONMENTS IS YOUR ORGANIZATION USING SYNTHETIC DATA TO PROTECT SENSITIVE DATA?

Application development and testing use cases are less popular with synthetic data users, with 63% and 62% using it in those use cases, respectively. Notably, **integration testing was a higher incidence use case with 74% of respondents citing it.**

Integration testing is an important use case whereby consistency of protection algorithm across systems is paramount. If you choose to use synthetic data for this use case, it is important to achieve accurate end-to-end testing and minimize integration errors in production. Therefore it is essential that your synthetic data solution provides data realism and referential integrity within and across data sources. It must mirror data relationships and complexity in your real production systems.

49% Use Synthetic Data in AI Development

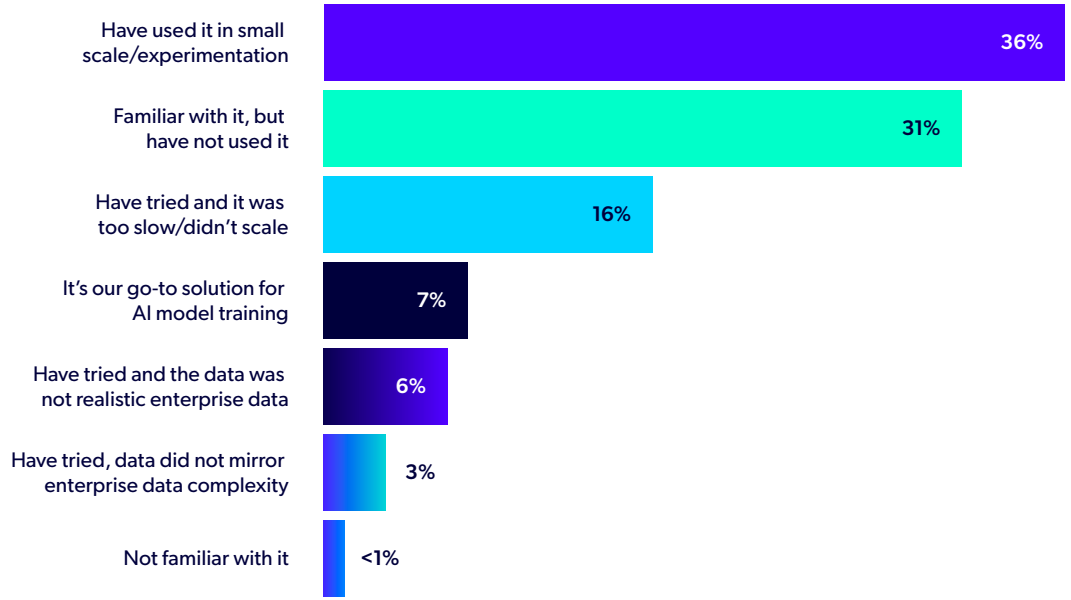
IN WHICH NON-PRODUCTION ENVIRONMENTS IS YOUR ORGANIZATION USING SYNTHETIC DATA?



We learned that 49% of those using synthetic data stated they use it for AI model training, fine-tuning, and RAG development. However, further probing revealed that only 7% use it as their go-to solution for AI model training.

In fact, **36% have only used synthetic data in small scale and experimentation mode**. And another **25% have tried it and experienced speed, scale, and data quality issues**.

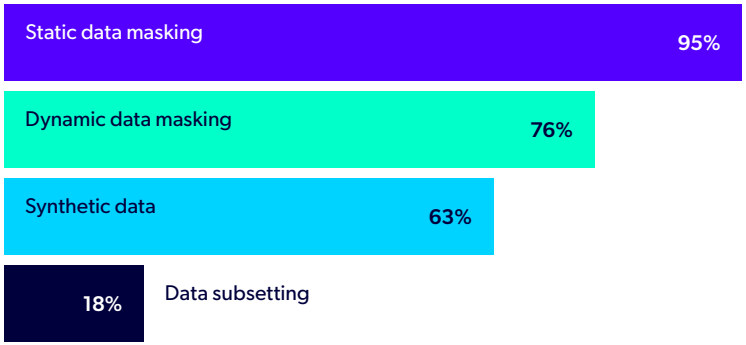
WHAT BEST DESCRIBES YOUR ORGANIZATION'S EXPERIENCE WITH THE USE OF SYNTHETIC DATA FOR AI MODEL TRAINING?



Enterprises Optimize with a Portfolio Approach to Compliance

Based on our findings, there is no "one size fits all" solution for protecting sensitive data. As we can see from the data, **organizations are clearly using more than one approach** — this was a "select all" question, and we can see the overlap in the data.

WHICH OF THE FOLLOWING DOES YOUR ORGANIZATION USE FOR PROTECTING SENSITIVE DATA IN NON-PRODUCTION ENVIRONMENTS? (E.G., SOFTWARE DEVELOPMENT, TESTING, AI)



While **static data masking is used by the vast majority of organizations (95%)**, it is by no means the only solution they use. **76% use dynamic data masking, 63% use synthetic data, and 28% use data subsetting**. Large enterprises commonly take a platform approach to data compliance and deploy a combination of use-case-fit approaches.

The Combined Approach

Static Masking for Compliance

All organizations we work with use static masking to permanently protect sensitive data in non-production environments when users don't need access to it. Static masking is applied when real data is available in production systems and can be masked before being provisioned to non-production environments. For example, in test data management for application development and testing, static masking is the preferred, safe, and compliant method to mitigate risks associated with test data.

DEFINITION
Irreversibly replaces sensitive data values with fictitious yet realistic equivalents that reflect production data and its complexity and relational attributes.

Dynamic Masking for Select Uses

Dynamic masking is ideally suited for production systems, such as a medical records system. A doctor needs to see a patient's PHI, but the finance team should not, so it is dynamically masked for that team.

DEFINITION
Role-based, temporary, reversible obfuscation, which may lead to data re-identification and breach by bad actors.

Dynamic data masking, common in production systems, sometimes extends to other use cases due to organizational inertia, even when a different approach is safer. Dynamic masking should never be used as the only defense in software development and testing environments; it leaves the underlying data unprotected and can lead to vulnerability. Conversely, static masking irreversibly protects data in those environments.

Dynamic masking may work well in analytics when executives need access to sensitive data in operational reports, while it's masked for other users. However, offshore teams developing BI reports should never see PII, and development data warehouses should be protected with static masking.

Synthetic Data for New Applications

Synthetic data is a valuable tool for enhancing development environments. Software teams can generate new custom data to test specific scenarios, such as user errors or unique behaviors, or use it for chaos testing.

DEFINITION
Synthetically-generated, new-new data that may or may not provide data realism and referential integrity vis-à-vis production data.

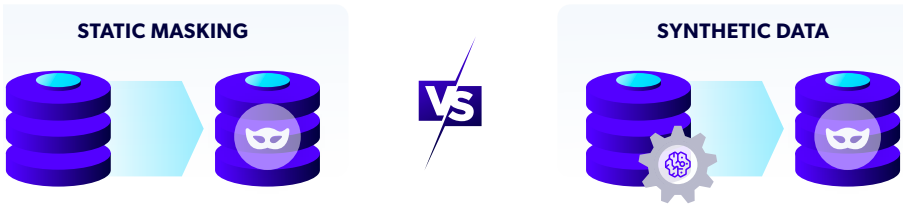
In some cases, teams may only need a basic dataset to ensure the software runs, where realistic masked data isn't necessary. For new applications with no production data, synthetic data becomes the only option.

When to Use Static Masking vs. Synthetic Data

In many cases, the best practice is to use static masking. However, some vendors add confusion by labeling this rule-based approach as synthetic data.

The exception is when an organization has a synthetic data mandate, prohibiting masked data in certain environments, such as offshore development.

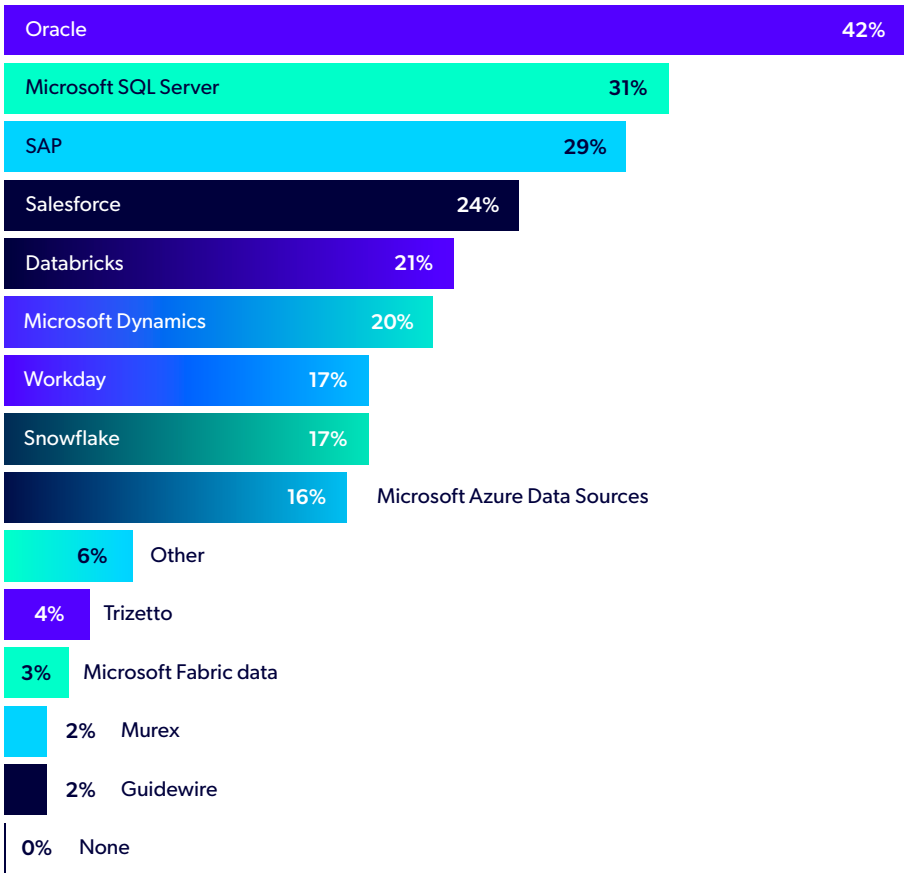
When evaluating data compliance approaches, consider key criteria like scale, speed, risk of exposure, referential integrity, data realism, and cost-effectiveness. For instance, for some use cases, you must ensure that synthetic data provides referential integrity and data realism.



Organizations Need to Protect Data Across Sources

Our research uncovered that organizations have broad charters for protecting sensitive data across the estate. This includes a wide variety of data sources that must be protected as well as multiple use cases, or non-production environment types, that require data compliance.

WHAT DATA SOURCES DOES YOUR ORGANIZATION NEED TO MASK? SELECT ALL THAT APPLY



Oracle and SQL Server are the Data Sources Organizations Need to Mask Most

42% of organizations currently protect sensitive data in non-production Oracle data sources. It is not surprising, given its wide prevalence in the market, that Oracle is the #1 most protected data system in non-production.

31% of organizations protect Microsoft SQL Server in non-production, making it the second most-protected data source. Again, this is not surprising considering the wide adoption of this data system.

Masking Off-the-Shelf Business Applications is Common

We learned that **29% of companies protect sensitive data in SAP non-production environments, 24% protect Salesforce, 20% protect Microsoft Dynamics, and 17% protect Workday.**

This data suggests that key third-party business applications are commonly protected and high on the list of priorities. Even vertical applications such as TriZetto (for healthcare insurance), Murex (for financial services), and Guidewire (for P&C insurance) had honorable mentions for masking priorities with 4%, 2%, and 2% respectively being masked in non-production.

We did not break down Oracle into key elements of the platform, so the 42% of Oracle masking could also represent EBS, to some extent.

Analytics & AI Sources are Also a Priority for Masking

21% protect Databricks in non-production and 17% protect Snowflake data. Given the rise of AI and the focus on protecting data in analytics environments, it is no surprise that Databricks is the fifth most popular source for masking — and that Snowflake made it into the top 10.

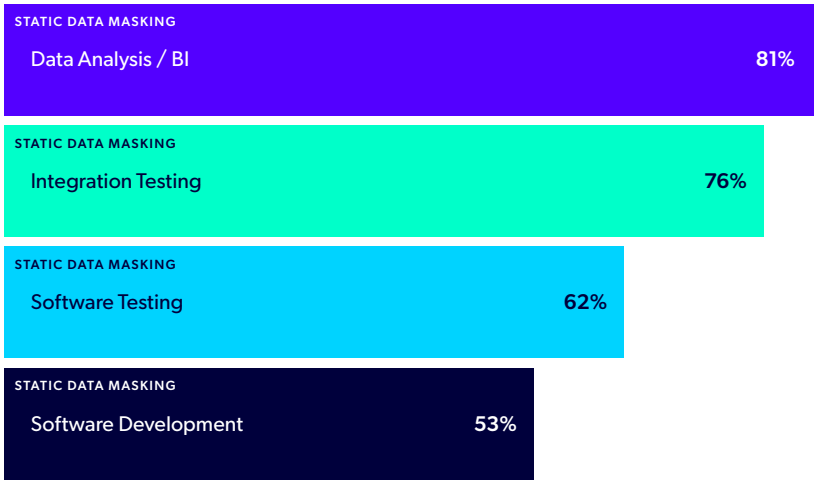
We were a bit surprised that masking Databricks beat Snowflake in popularity, given the relative market prevalence of these sources. But the fact Databricks has become associated with the mad dash to adopt AI could be one explanation.

We are getting many inquiries related to data compliance in both of these sources and fully anticipate the popularity of protecting these two and other AI and analytics data sources will only increase.

Analytics is the #1 Use Case for Protecting Sensitive Data

We found out that **81% of organizations that use static data masking are using it to protect sensitive data in analytics and BI environments**. Analytics was also the number #1 use case for those using dynamic data masking, synthetic data, and data subsetting.

IN WHICH NON-PRODUCTION ENVIRONMENTS IS YOUR ORGANIZATION USING STATIC DATA MASKING TO PROTECT SENSITIVE DATA?



This is also in line with our earlier finding that analytics environments are the top reason for the increase in the sensitive data footprint in non-production (according to 65% of respondents).

Clearly, with the increased use of data for decision making, there is also an increased need to protect that data.

The Need to Protect Sensitive Data in Software Development and Testing is High

76% of those using static data masking apply it to protect sensitive data in integration testing. This also aligns to the high incidence of protecting sensitive data in third-party applications such as SAP, Salesforce, and Dynamics, cited earlier. These applications are commonly used together in end-to-end business workflows such as "order to cash" and therefore are included in integration testing.

Two considerations in choosing an approach for protecting sensitive data in integration testing:

1. Ensure referential integrity across data sources.
2. Avoid using only dynamic masking in integration testing, as the underlying data remains exposed and there is no need for those involved in integration testing to see PII. Similarly, row subsetting leaves the organization exposed, as it still contains sensitive data.

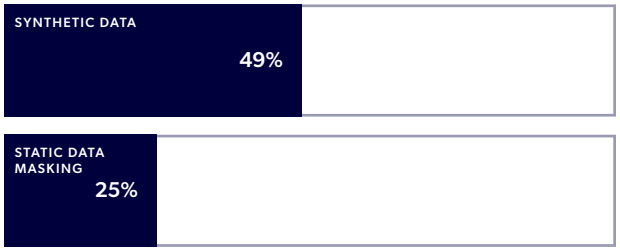
The **third and fourth most common use cases for static data masking were software testing and software development, with 62% and 53% respectively** being cited by organizations. Software development and testing are considered a security "weak link" of organizations and one of the biggest threat vectors from a cybersecurity standpoint.

It is not surprising that so many organizations want to protect those environments. We would like to see that trend increase. And we discourage the use of dynamic masking and subsetting for these use cases, for the same reason as above.

Protecting Sensitive Data in AI is an Emerging Trend

49% of those using synthetic data are using it to avoid the introduction of sensitive data in AI environments. There is also adoption of static data masking as a means to protect sensitive data; 25% of those using static masking reported using it for this purpose.

WHICH DATA MASKING METHODS ARE YOU USING TO PROTECT SENSITIVE DATA FOR AI MODEL TRAINING / FINE-TUNING AND RAG DEVELOPMENT?



As we discussed earlier, the distinctions between synthetic data and static data masking are nuanced and vary by your definition of synthetic data. As the market for AI and data privacy matures, we will gain greater clarity on which approach is best suited for specific use cases.

There's Mass Confusion About the Exposure of Sensitive Data in AI Model Training and Fine-Tuning

In this year's survey, we went deeper into enterprise perceptions of AI and data privacy as this disruptive technology becomes more mature. What we found was… tremendous confusion.

To start, we found that **94% of organizations are already past the initial stages of AI adoption**.

- 61% are standardizing AI — meaning formalizing AI usage with dedicated resources and training.
- 14% are transforming with AI — meaning AI is deeply integrated into business processes and drives competitive advantage.

AI and analytics workflows rely heavily on data, and as we can see, organizations are moving up in the AI maturity curve. With that, they are pushing teams to get the data they need as fast as possible.

Our findings indicate that organizations overlook the risk of using sensitive data in their rush to execute on AI. They also tend to believe that they need to use real data — because AI models perform best when trained on large quantities of realistic data.

But we found some surprising contradictions in the responses:

- **91% say sensitive data should be allowed in AI training and testing.**
- **82% believe it is safe to do so.**
- **Yet, 78% are highly concerned about theft or breach of model training data.**
- **And 68% worry about privacy and compliance audits.**

Advice from the Authors

If we can offer one piece of definitive guidance: **you should never use real, identifiable data to train AI models**. Therefore, the use of row subsetting to train and fine-tune AI models is highly discouraged, and column subsetting may cause the model to miss important characteristics of the data. Dynamic masking can also be risky, depending on how exactly it is used in the context of AI model training and fine-tuning, since it is reversible. So, proceed with caution, and guard your sensitive data to ensure secure and responsible AI innovation.

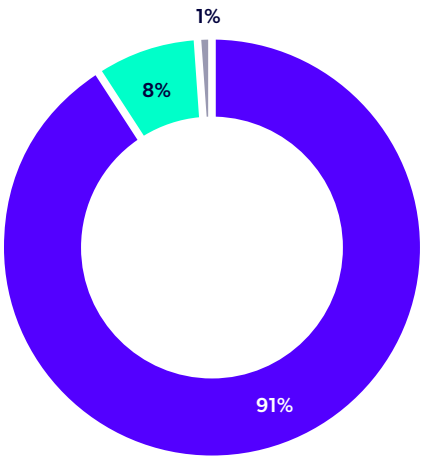
Let's expand on each of these results, then explore the apparent confusion in the market about the safety of using sensitive data in AI environments.

The Majority Want to Allow Sensitive Data in AI — And Think it's Safe

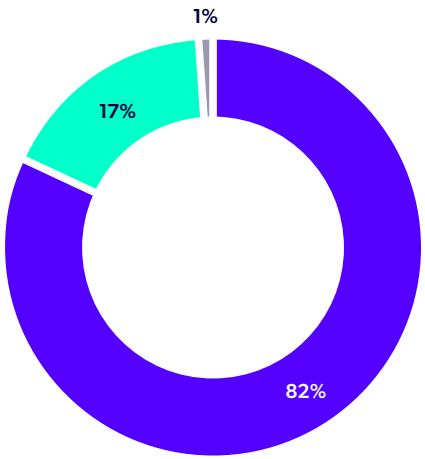
An astounding **91% of respondents said sensitive data should be allowed in AI model training, fine-tuning, and RAG (retrieval augmented generation) application development and testing.**

Fewer respondents (82%), but still most, say they think it is safe to use sensitive data in AI model training and fine-tuning. **Only 17% said it isn't safe**.

DO YOU BELIEVE SENSITIVE DATA (PII/PHI) SHOULD BE ALLOWED IN AI MODEL TRAINING, FINE-TUNING, AND RETRIEVAL AUGMENTED GENERATION (RAG) APPLICATION DEVELOPMENT/TESTING?



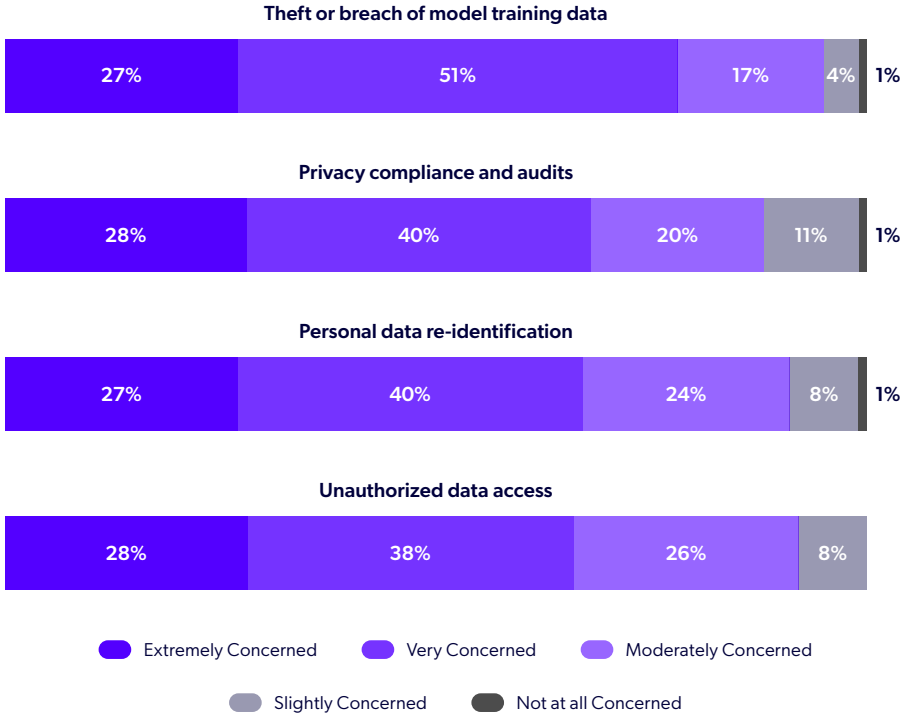
DO YOU BELIEVE IT IS SAFE TO USE SENSITIVE DATA IN AI MODEL TRAINING AND FINE-TUNING?



Yet, Most are Concerned About the Security of AI Training Data

Our results indicate that, while leaders want to use sensitive data in AI training, and they think it's safe, they are at the same time quite concerned about the security of doing so. **Between 88% and 95% of respondents are at least moderately concerned about theft or breach of model training data, privacy compliance and audits, personal data re-identification, and unauthorized data access in AI development and training environments**.

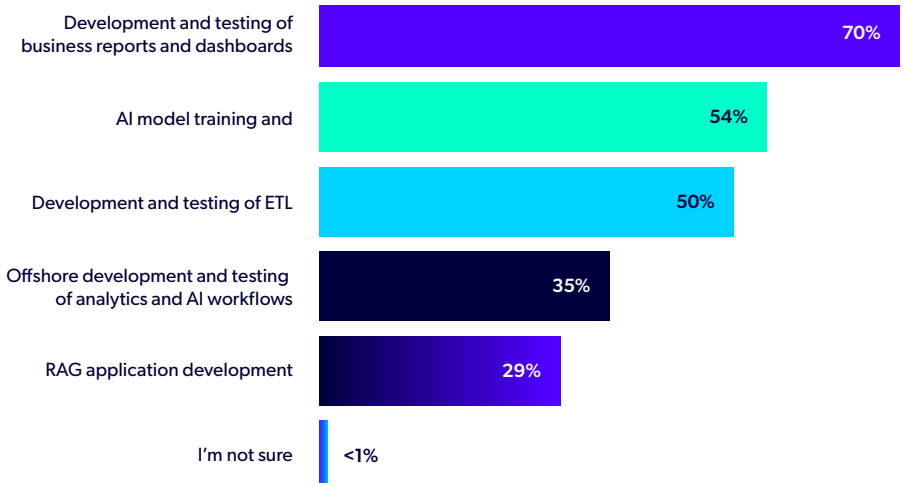
WHAT IS YOUR LEVEL OF CONCERN WITH THE FOLLOWING INVOLVING AI DEVELOPMENT AND TRAINING ENVIRONMENTS AT YOUR ORGANIZATION?



Most Need to Protect Sensitive Data in Analytics & AI

Additionally, **a large majority (70%) say they need to protect sensitive data in non-production to support analytics**, i.e., the development and testing of business reports and dashboards. And **more than half (54%) say they need to protect it in AI model training and tuning**.

DO YOU NEED TO PROTECT SENSITIVE DATA IN A NON-PRODUCTION VERSION OF YOUR DATA WAREHOUSE OR DATA LAKE IN ORDER TO SUPPORT ANY OF THESE USE CASES?



Analytics

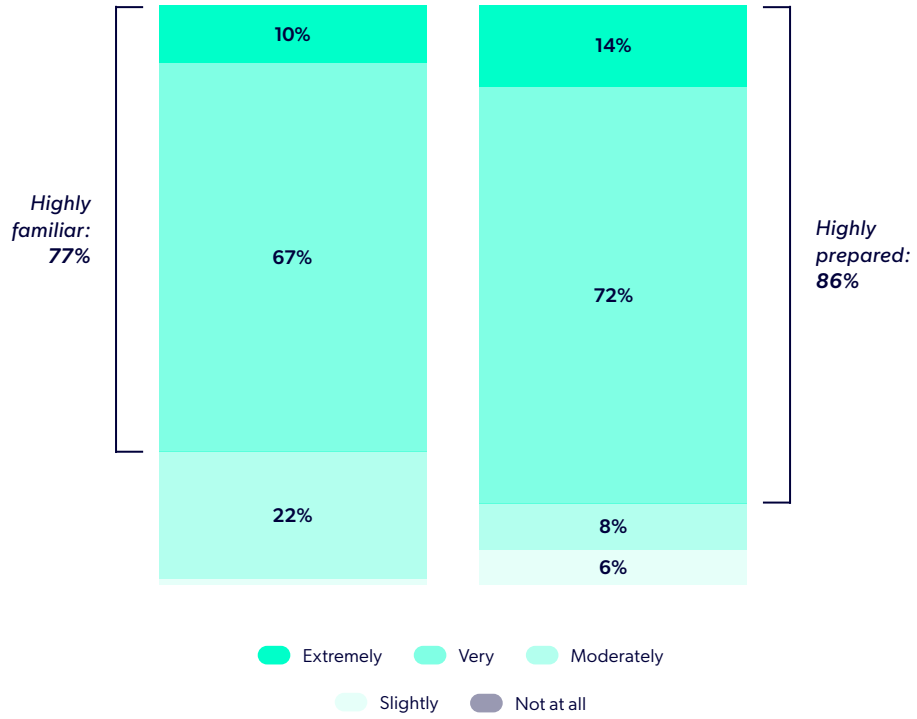
Organizations prioritize protecting sensitive data for analytics. Platforms like Databricks and Snowflake are key analytics sources that need to be protected, with 21% and 17% of respondents saying they need to mask them, respectively. (See chart on page 23.)

Most Know Where Sensitive Data is Being Used in AI Development

Adding to the complexity of our findings, 77% of respondents reported being either highly or extremely familiar with where sensitive data is currently used in AI development. And even more — **86% — feel that their organization is very or extremely prepared for additional or expanded AI regulations**.

HOW FAMILIAR ARE YOU WITH WHERE SENSITIVE DATA IS CURRENTLY USED IN AI DEVELOPMENT?

HOW PREPARED IS YOUR ORGANIZATION FOR ADDITIONAL OR EXPANDED AI REGULATIONS?
(E.G., EU AI ACT, APPLYING EXISTING REGULATIONS LIKE GDPR FOR AI, ETC.)



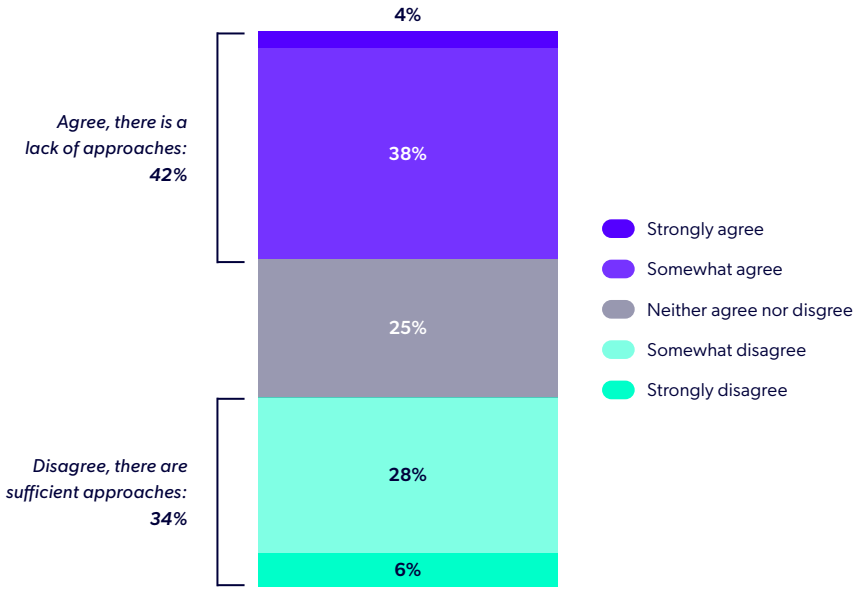
Less Than Half Believe They Have Sufficient Solutions for Data Privacy in AI

We asked respondents how much they agree with the statement, "I believe there is a lack of approaches and tools to tackle data privacy in AI environments." This year, less than half agreed there is a lack of approaches (42%), compared to last year when 64% agreed — a decrease of 34%.

Only **34%** believe there are sufficient approaches and tools for tackling data privacy in AI environments.

PLEASE INDICATE HOW MUCH DO YOU AGREE OR DISAGREE WITH THE FOLLOWING STATEMENT:

"I believe there is a lack of approaches and tools to tackle data privacy in AI Environments"



Much can change in a year, especially as new solutions emerge, and we are glad to see that more organizations believe there are solutions for protecting data privacy in AI.

As for the 25% who neither agreed nor disagreed? They're confused. More on this later.

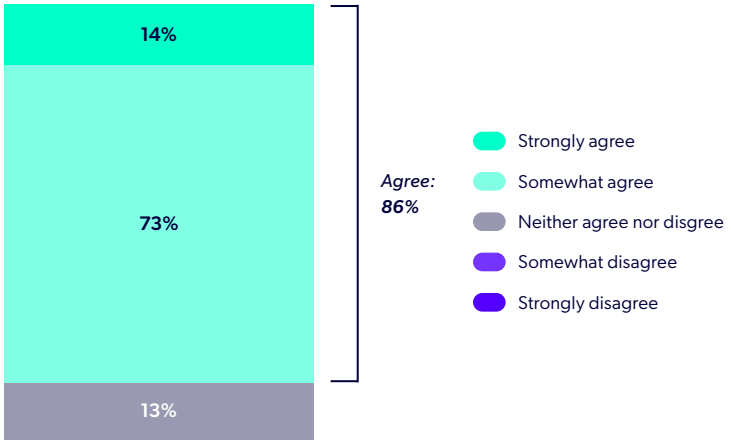
Most Organizations Plan to Invest in AI Data Privacy Immediately

86% agreed with the statement, "My organization is looking to invest in AI data privacy in 2025-2026." So, data privacy is clearly a key priority for organizations as they mature in AI and advance development in AI environments.

86% plan to invest in AI data privacy in 2025-2026

PLEASE INDICATE HOW MUCH DO YOU AGREE OR DISAGREE WITH THE FOLLOWING STATEMENT:

"My organization is looking to invest in AI data privacy in 2025 - 2026"



"The fact that 86% of organizations plan to invest in AI and data privacy highlights the growing recognition that innovation and governance must go hand in hand. As AI capabilities expand, the value of trusted, well-managed data becomes even more critical. Success will come to those who can harness AI to drive insights while ensuring that privacy, security, and compliance remain foundational to every initiative."

— Azhar Rahman,
Enterprise Data Director
AgFirst Farm Credit Bank

There's a Clear Disconnect When it Comes to AI and Data Privacy

The tension between organizations' ambitions with AI and their concerns about sensitive data underscores a critical disconnect in the market.

Contradictory Sentiments About Sensitive Data Use

On one hand, as we saw, the vast majority **(91%) of respondents think sensitive data should be allowed in AI model training and testing**, with 82% believing it is safe to do so. Simultaneously, a significant majority (**78%) express deep concern about risks like theft or breach of training data**.

These contradictions suggest that there is a lack of shared understanding or consistent guidance within organizations about the safety of using sensitive data in AI model training and initiatives. They may not be aligned on how to balance the promise of AI with the risks posed by sensitive data exposure.

Rapid Adoption Outpaces Security Frameworks

Part of this confusion may stem from the rapid pace of AI adoption. AI's spread has outpaced the development of comprehensive frameworks for managing sensitive data in its environments. Many organizations are venturing into uncharted territory. They are leveraging sensitive data to improve AI performance without the deployment of robust data security and compliance solutions.

Gaps in Tools and Approaches for Data Privacy

Compounding the problem is the evident lack of sufficient tools and approaches specific to data privacy in AI. While 77% of respondents feel highly familiar with where sensitive data is used in AI environments, only 42% believe there are adequate solutions to ensure data privacy. This contradiction between the confidence organizations have about their current exposure and the uncertainty they feel about their present solutions is interesting.

Building a Secure Foundation for Innovating with AI

The "AI arms race" is pushing teams to adopt AI as fast as possible, and as a result, they are pushing back on any extra processes that may slow them down or hurt the overall quality of their AI initiatives. At the same time, the majority of leaders are extremely concerned about AI data privacy.

These agendas — move fast with AI, and/or keep customer and proprietary data safe — seem to compete.

It's true that to safely unlock AI's potential, organizations must address the foundational challenges of securing sensitive data in AI workflows. AI models rely on large amounts of data to deliver accurate results, which is why so many organizations feel pressured to use real customer data during training and testing. But using sensitive data in AI poses tremendous risks, including regulatory violations, data reidentification, breach and theft, and unpredictable resharing of data by models. All of these can cause reputational damage and financial loss.

Seeing as more than two thirds of respondents expressed concern about theft of training data (95%), compliance and audits (88%), and personal data re-identification (91%) with regards to AI development, it's clear that most leaders are aware of the potential risks. But they likely don't understand them, and they do not seem to know how to address them.

One key solution: By permanently replacing sensitive information with realistic but synthetic data in non-production environments, organizations can train and test AI models without compromising their customers' privacy. This ensures that datasets remain secure and maintain their structure and utility for accurate AI performance.

When combined with robust privacy frameworks, using statically masked data — and, optionally, synthetic data generation for unknown or less-important data — lets businesses harness AI's potential while safeguarding data and meeting compliance requirements.

"The disconnect between confidence in using sensitive data for AI and the deep concern over its potential breach highlights a critical trust gap. As we scale AI adoption, we need to prioritize privacy-preserving architectures and responsible data practices that go beyond compliance. Sustainable innovation means building systems that are not only powerful, but also secure, ethical, and resilient by design."

rackspace technology

— Ben Blanquera,
VP of Technology and Sustainability,
Rackspace

Get a deeper dive on the risks AI poses to data privacy and compliance + best practices to implement now in our expert guide, "*AI Without Compromise*."

Get AI Data Privacy Guidance

Accelerate Data Compliance with Perforce Delphix

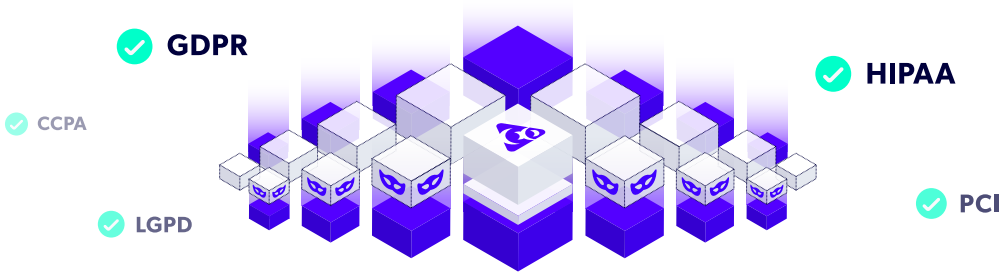
Sensitive data sprawls from production to non-production environments, including software development, testing, analytics, and AI. This data is governed by expanding privacy regulations such as GDPR and HIPAA, increasing organizational exposure and risk.

Protecting this data is a critical priority for enterprises. However, manual or native data masking methods can create bottlenecks — especially in complex, large-scale estates.

Perforce Delphix empowers you to ensure data compliance, accelerate development speed, and increase quality, without trade-offs, through a combination of static data masking and AI-powered synthetic data capabilities integrated into your DevOps and AI pipelines.

That's because Delphix:

- Eliminates sensitive data risks.
- Prevents data theft.
- Ensures regulatory compliance.
- Accelerates innovation.
- Assures software, analytics, and AI quality.



Protect Sensitive Data

With Delphix, 77.2% more data and data environments are masked and protected.* Delphix combines static data masking and AI-powered synthetic data capabilities to automatically discover and replace sensitive information with realistic, production-like values. Benefit from automated sensitive data discovery and a robust library of pre-built, customizable algorithms to maintain security, data utility, and referential integrity across all sources — from on-premises to cloud.

Gain Speed At Enterprise Scale

Delphix scales from the smallest database to massive, multi-billion row analytical platforms like Snowflake and Databricks. It rapidly delivers compliant data to downstream teams, when and where it's needed. Integration with DevOps and AI pipelines enables you to shift left and boost quality.

Achieve Compliance — Without Trade-Offs

With Delphix, achieve data compliance while increasing speed and quality of development, analytics, and AI initiatives. No trade-offs required.

Request your demo of Delphix today to discover how AI-powered compliance can help secure your data and accelerate your business.

Request a Compliance Demo ▶

perforce.com/products/delphix/demo/compliance

*IDC Business Value White Paper, sponsored by Delphix, by Perforce, The Business Value of Delphix, #US52560824, December 2024

DELPHIX IS TRUSTED BY

3 of the top 5

Financial Institutions

4 of the top 5

Healthcare Insurers

Gartner

"The integration of the virtualization and masking solutions makes it simple for us to consistently protect sensitive data in our pre-production environments. The self-service capabilities allow our developers to perform testing more efficiently."

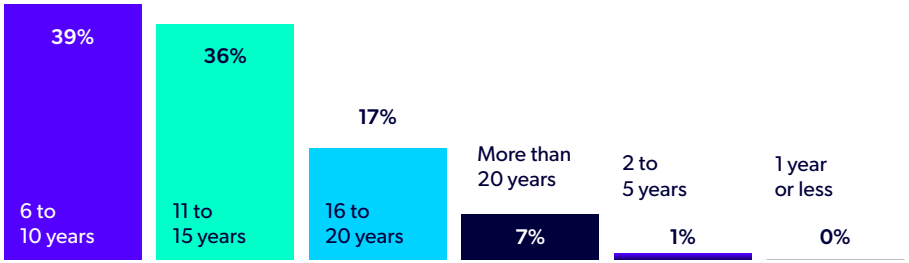
Gartner Peer Insights Customers' Choice 2025

IT User, Insurance Industry
Gartner Peer Insights ™

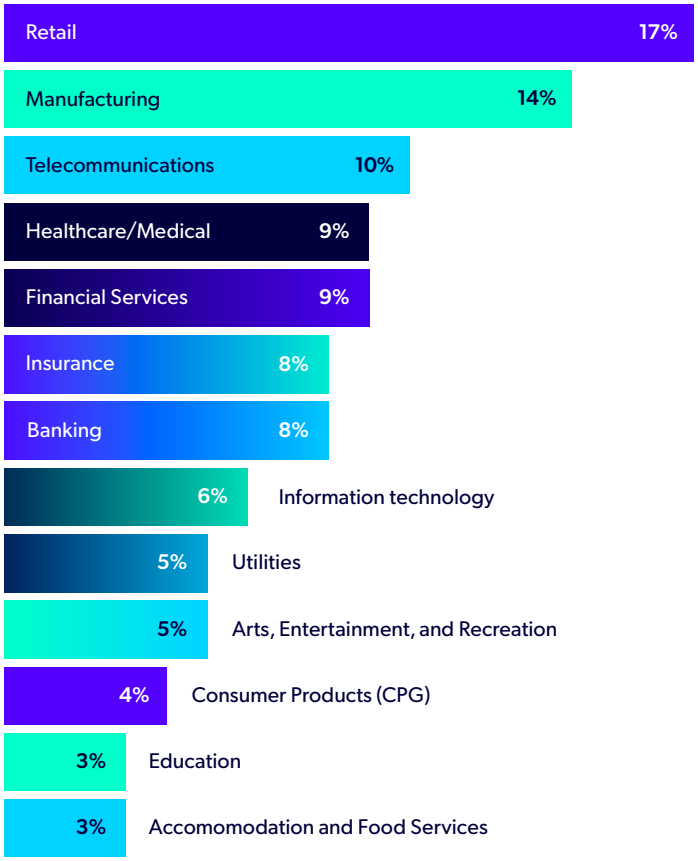
Full Audience Demographics

Perforce Delphix partnered with a third-party research firm to survey 280 enterprise leaders across the globe. The data collected is the basis for this report.

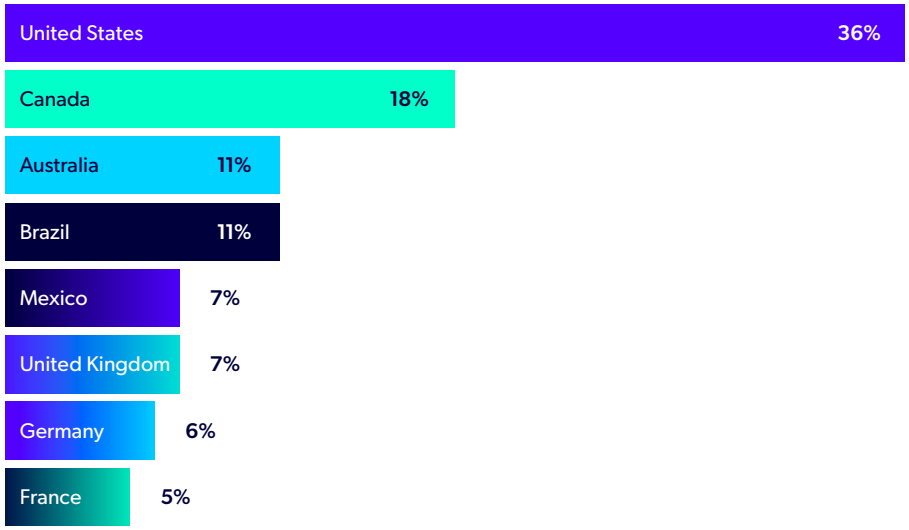
YEARS OF EXPERIENCE



INDUSTRY

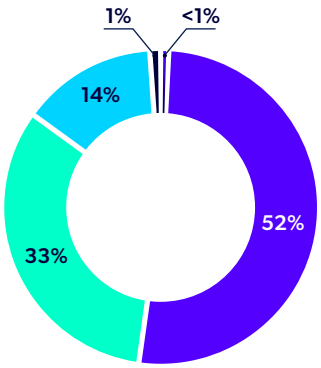


COUNTRY



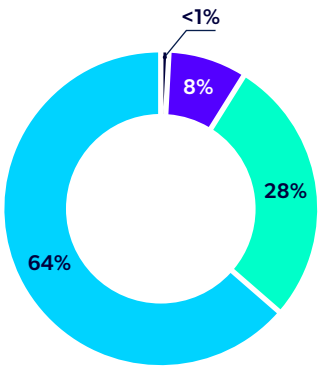
DECISION-MAKING STATUS

JOB ROLE



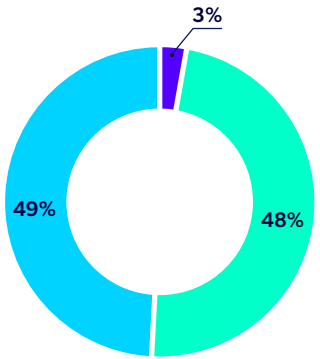
- Manager/Sr.Manager
- Director
- VP/Senior VP
- C-Suite Executive
- Technology Architect

HOW OVERALL BUDGET CHANGED FROM 2024 TO 2025



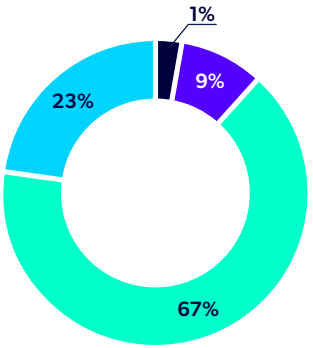
- Decreased
- Stayed the same
- Increased
- I'm not sure

FREQUENCY OF WORKING WITH SENSITIVE CONSUMER DATA



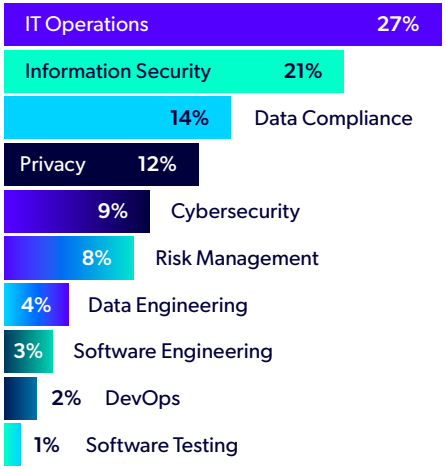
- Sometimes
- Often
- Very Often

AGE

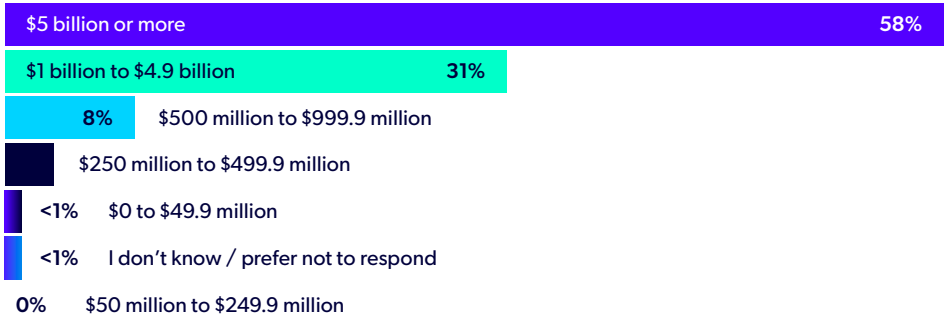


- 25 to 34
- 35 to 44
- 45 to 54
- 55 +

JOB FUNCTION



ORGANIZATION ANNUAL REVENUES



Key Terms

Data Masking: The process of protecting sensitive data by replacing real values with fictitious values.

Dynamic Data Masking: Role-based, temporary, reversible obfuscation, which may lead to data re-identification and breach by bad actors.

Expanded Exposure Footprint: The increased sensitive data risk or vulnerability that can occur in non-production environments due to potential security gaps and the expectation that each production environment often has many corresponding non-production environments in a typical enterprise. For example, a single production application may have 7–10 non-production environments used for development and testing. If these environments are not protected properly, they can become an easy target for cyberattacks and create expanded regulatory exposure.

Live Environment: See production environment.

Lower Environment: See non-production environment.

Non-Production Environment: Any environment not intended for production (also known as live) use. Common examples of non-production environments include development, testing, analytics, and AI/ML environments. Non-production environments are also known as lower environments.

Personal Identifiable Information (PII): Any data that can identify a person.

Production Environment: The live environment that contains the latest version of the application, software, or product.

Production-Like Data: See realistic data.

Protected Health Information (PHI): Any data that can identify a patient.

Realistic Data: Masked data that resembles real data. This is also known as production-like data.

Sensitive Data: Information that must be protected and kept confidential. Examples include sensitive consumer data, such as PII and PHI.

Static Data Masking: Permanently replaces sensitive information, such as PII and PHI with fictitious, yet realistic data.

Data subsetting: A minimized production dataset that redacts table columns for sensitive fields.

Synthetic Data: Synthetically-generated, net-new data that may or may not provide data realism and referential integrity vis-a-vis production data.

Unrealistic Data: Fake data that does not maintain the appearance and usability of real data.

About the Authors

Ann Rosen

Director of Product Marketing, Perforce Delphix

Ann is focused on data compliance and protecting sensitive data across software development and testing as well as AI and analytics environments, without compromising innovation, speed, and quality. She also focuses on DevOps test data management to securely accelerate digital transformation. Ann is passionate about bringing positive change through technology and has been involved in enterprise software and services for over 20 years. She has specialized in areas such as cloud data management and protection, Java, and software development. Prior to Delphix, Ann led product marketing in organizations of various sizes and maturity, including Informatica, Druva, and Sun Microsystems.

Steve Karam

Principal Product Manager, Perforce Delphix

Steve Karam is an enterprise technologist and outcome-driven leader with diverse expertise in customer success, product management and engineering, education, SaaS, cloud, data technologies including AI/ML and natural language processing, and beyond across nearly every market vertical. A restless and relentless learner and advocate for personal and team growth who works primarily with federal, financial services, and healthcare organizations within Perforce. In the technical arena Steve is proficient or expert level with cloud architecture, multiple RDBMS, NoSQL, multiple programming languages including Python and React, common DevOps & SDLC tools, AI/ML, SaaS architecture, blockchain, data analytics, and several other topics.

Ross Millenacker

Senior Product Manager, Perforce Delphix

Ross Millenacker leads the Perforce Delphix compliance portfolio, overseeing Continuous Compliance and Hyperscale Masking. He focuses on advancing core compliance capabilities and delivering innovative solutions to help customers manage sensitive data at scale. Before joining Delphix, Ross held roles spanning Product Marketing and Engineering in the software and semiconductor industries. He holds an MBA from Carnegie Mellon, a BS in Mechanical Engineering from UC Berkeley, and a BA in History from UC Santa Cruz.

Contributing Editor

Skye Pinney

Skye Pinney is a Content Marketing Manager for Perforce Delphix. With a background in professional writing and rhetoric, she has spent her marketing career transforming complex technical concepts and disparate messaging into clear, compelling brand stories. She loves to create content that educates, inspires, and builds trust.

About Delphix

Perforce Delphix is the industry leader in automated, compliant data delivery, enabling enterprises to drive DevOps agility, AI innovation, and cloud transformation. By removing traditional bottlenecks, our platform allows organizations to release applications 2x faster, secure sensitive data with 77% less exposure, and reduce infrastructure costs by 80%.

With advanced data virtualization, automated data masking, ephemeral cloud data environments, and integrations with DevOps and AI tools, Delphix ensures data is instantly accessible, secure, and compliant. Global leaders like Dell, BNP Paribas, Michelin, and Morgan Stanley rely on Delphix to fuel innovation, strengthen compliance, and optimize IT efficiency.

Request Demo

perforce.com/products/delphix/demo/compliance

About Perforce

The best-run DevOps teams in the world choose Perforce. Powered by advanced technology, including powerful AI that takes you from AI ambition to real results, the Perforce suite is purpose-built to handle complexity, maintain speed without compromise, and ensure end-to-end integrity across your DevOps toolchain. With a global footprint spanning more than 80 countries and including over 75% of the Fortune 100, Perforce is the trusted partner for innovation.

Harness the power of AI and accelerate your technology delivery without shortcuts. Build, scale, and innovate with Perforce—where efficiency meets intelligence.

Contact Us

perforce.com/contact-us