perforce
Delphix

REPORT

2025 State of Data Masking Solutions Report

Excerpted from the **2025 State of Data Compliance and Security Report**



© Perforce Software, Inc. All trademarks and registered trademarks are the property of their respective owners. (0920NL25)

Contents

3..... **A Letter From the Authors**

4 ..... **Who Took the Survey?**

5..... 95% of Organizations Use **Static Masking**

8 ..... 63% of Organizations are Using **Synthetic Data**

10 ..... Enterprises Optimize with a **Portfolio Approach to Compliance**

15 ..... Secure Your Data with Perforce **Delphix Data Masking**

© Perforce Software, Inc. All trademarks and registered trademarks are the property of their respective owners. (0920NL25)



Ann Rosen
Director of Product Marketing



Steve Karam
Principal Product Manager



Ross Millenacker
Senior Product Manager

A Letter From the Authors

Welcome to this special report on the state of data masking solutions today, an excerpt from our comprehensive 2025 State of Data Compliance and Security Report.

Data compliance and security in non-production environments remain a critical priority for enterprise success. Our research confirms that most organizations use multiple approaches to protecting sensitive data, with static masking being the most popular option. We found that 95% of organizations use static data masking to irreversibly protect sensitive data, and 81% find it highly effective for securing data analytics and BI environments.

We also found that dynamic masking and synthetic data are the next most popular methods. Large enterprises commonly take a platform approach to data compliance and deploy a combination of use-case-fit approaches.

In this focused report, we dive into more data on how enterprise teams use different masking types, plus offer guidance on what each type is best suited for.

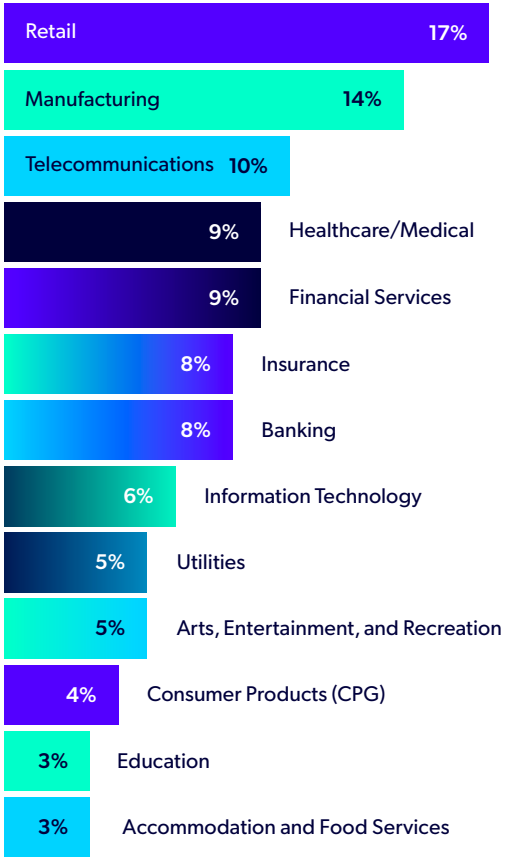
Sincerely,

Ann Rosen, Steve Karam, Ross Millenacker

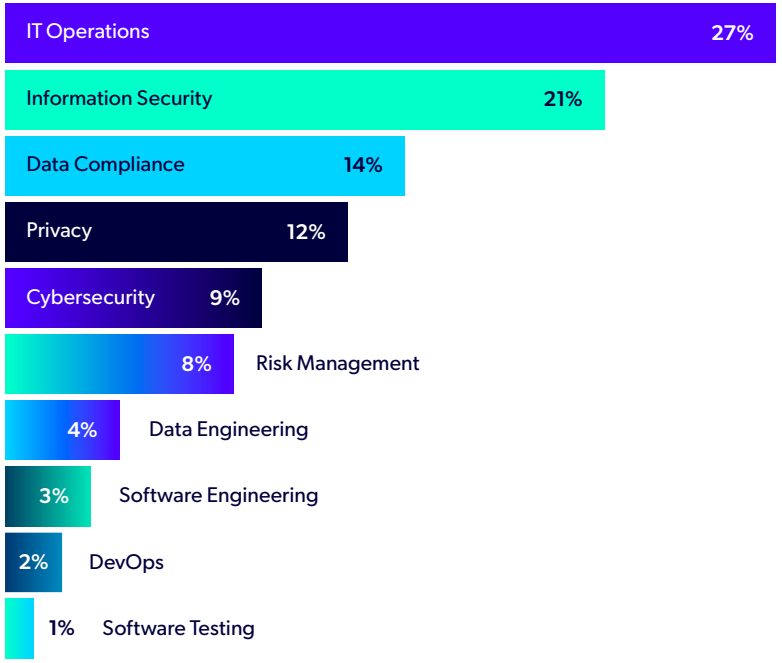
© Perforce Software, Inc. All trademarks and registered trademarks are the property of their respective owners. (0920NL25)

Who Took the Survey?

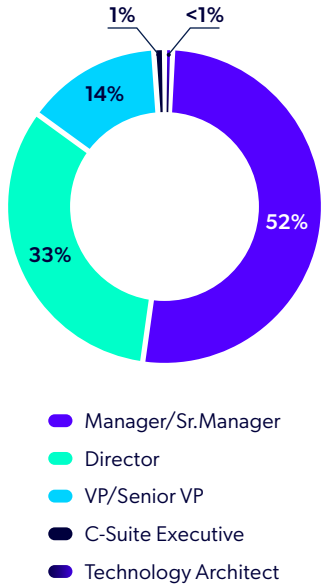
INDUSTRY



JOB FUNCTION



JOB ROLE



This report is excerpted from the 2025 State of Data Compliance and Security Report. For full demographics plus further research, download the full report.

© Perforce Software, Inc. All trademarks and registered trademarks are the property of their respective owners. (0920NL25)

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

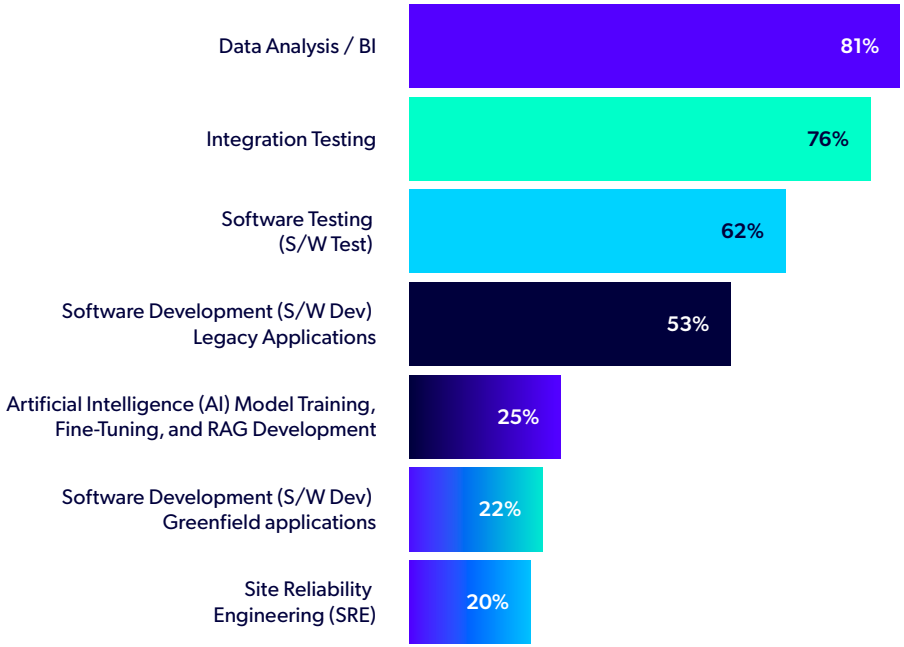
© Perforce Software, Inc. All trademarks and registered trademarks are the property of their respective owners. (0920NL25)

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.

© Perforce Software, Inc. All trademarks and registered trademarks are the property of their respective owners. (0920NL25)

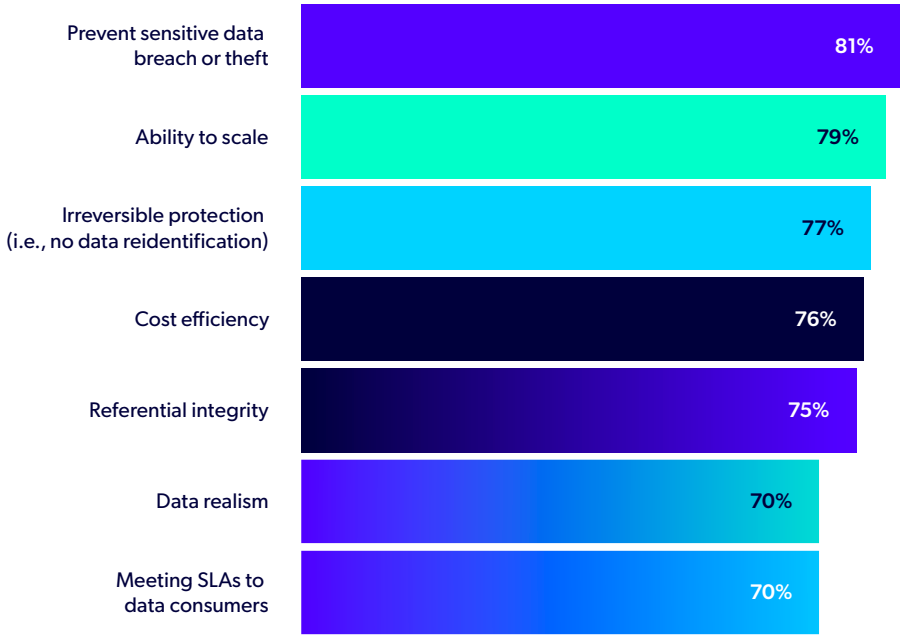
Key Qualities of Static Masking

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

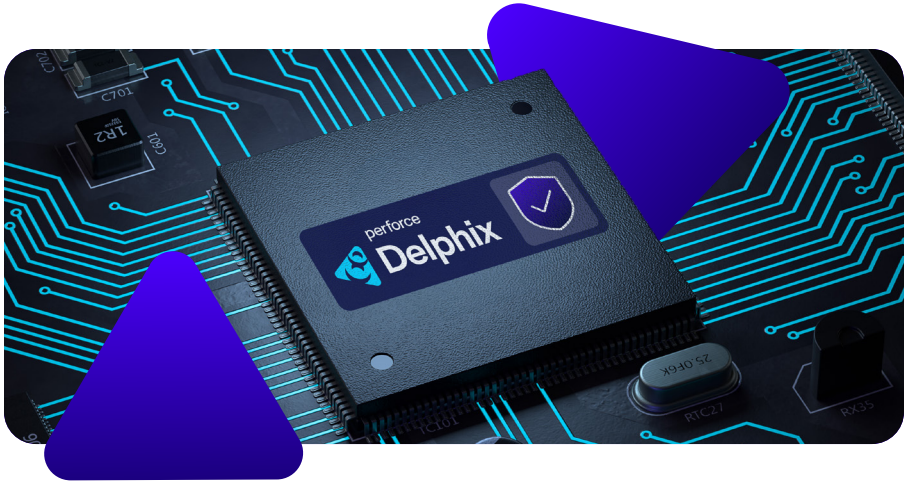
When implemented correctly, static masking can scale across large, complex enterprise data environments. High-performance techniques like parallel processing and automation, combined with efficient delivery methods such as data virtualization, can drastically reduce the time to deliver compliant data — 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.

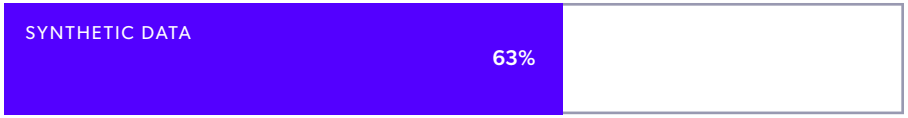
© Perforce Software, Inc. All trademarks and registered trademarks are the property of their respective owners. (0920NL25)



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.

© Perforce Software, Inc. All trademarks and registered trademarks are the property of their respective owners. (0920NL25)

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.**

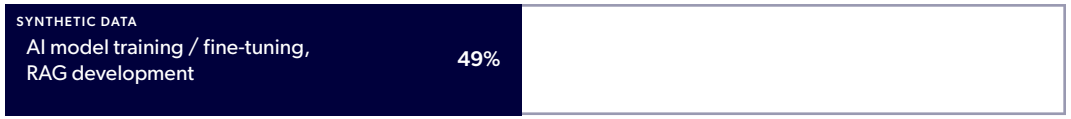
Integration testing is an important use case whereby consistency of protection algorithm across systems is paramount. If you choose to use synthetic data for this use case, it is important to achieve accurate end-to-end testing and minimize integration errors in production. Therefore it is essential 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.



© Perforce Software, Inc. All trademarks and registered trademarks are the property of their respective owners. (0920NL25)

49% Use Synthetic Data in AI Development

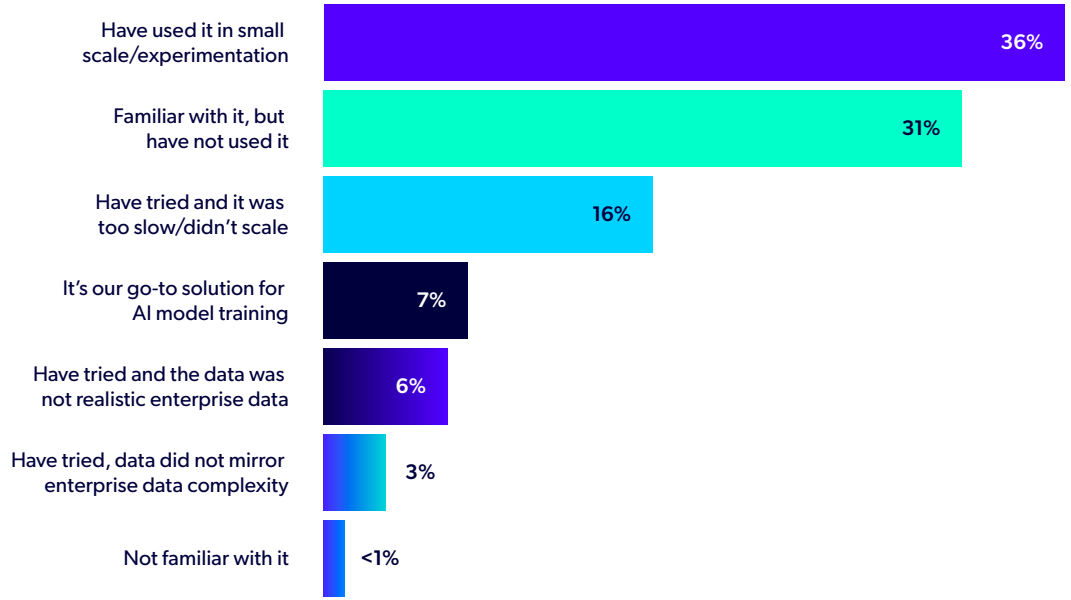
IN WHICH NON-PRODUCTION ENVIRONMENTS IS YOUR ORGANIZATION USING SYNTHETIC DATA?



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**.

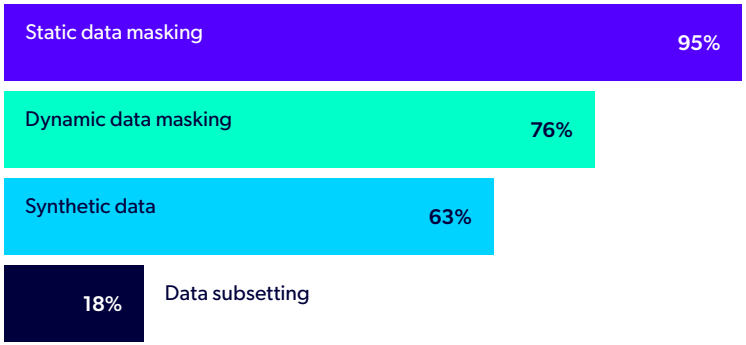
WHAT BEST DESCRIBES YOUR ORGANIZATION'S EXPERIENCE WITH THE USE OF SYNTHETIC DATA FOR AI MODEL TRAINING?



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.

© Perforce Software, Inc. All trademarks and registered trademarks are the property of their respective owners. (0920NL25)

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.

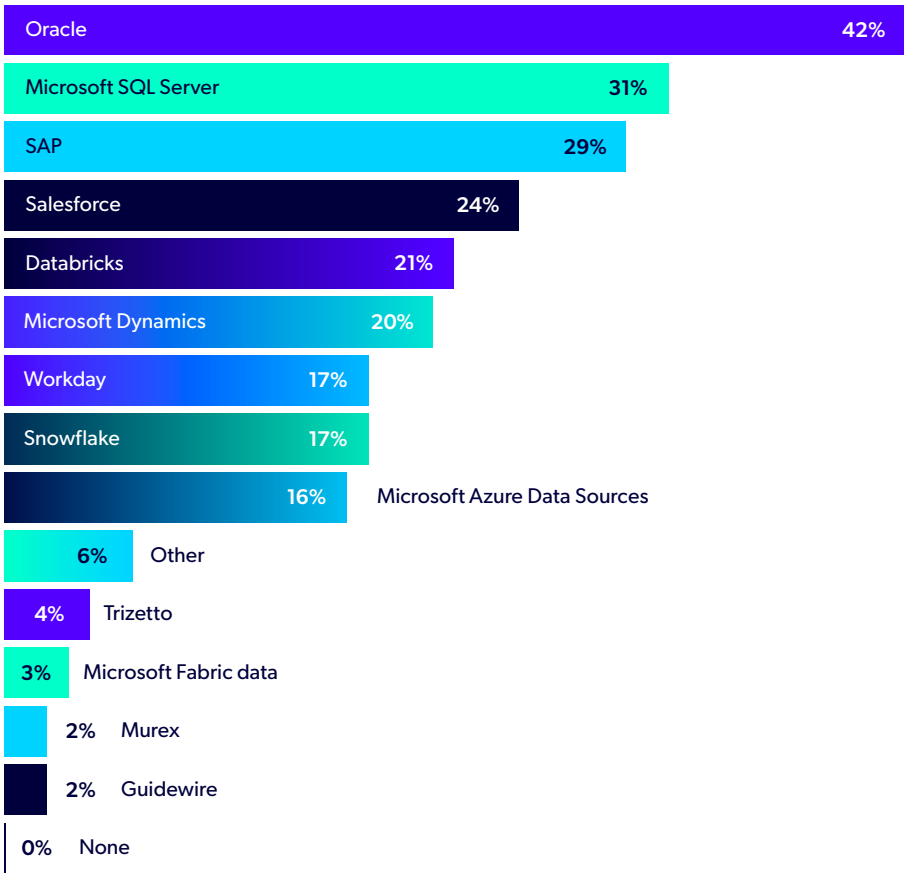


© Perforce Software, Inc. All trademarks and registered trademarks are the property of their respective owners. (0920NL25)

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.

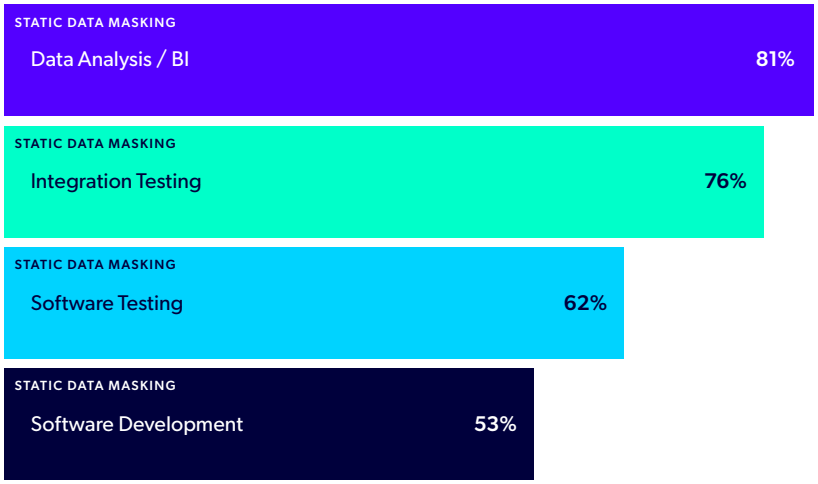
© Perforce Software, Inc. All trademarks and registered trademarks are the property of their respective owners. (0920NL25)

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.

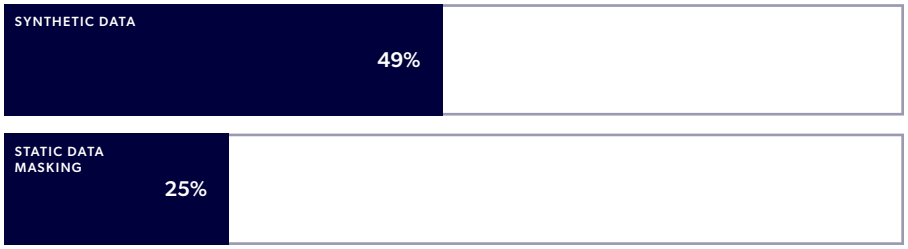
It is not surprising that so many organizations want to protect those environments. We would like to see that trend increase. And we discourage the use of dynamic masking and subsetting for these use cases, for the same reason as above.

© Perforce Software, Inc. All trademarks and registered trademarks are the property of their respective owners. (0920NL25)

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.

© Perforce Software, Inc. All trademarks and registered trademarks are the property of their respective owners. (0920NL25)

Secure Your Data with Perforce Delphix Data Masking

Allowing sensitive data to proliferate across non-production environments is a significant compliance and security risk. Perforce Delphix delivers powerful, automated data masking to protect this data across your enterprise. By replacing sensitive data with fictitious yet realistic values, Delphix ensures you can innovate securely without compromising speed or quality. With Delphix, 77.2% more data and data environments are masked and protected.*

Get Automated, Scalable Masking

Delphix automatically discovers and consistently masks sensitive data across your entire data estate. Easily define masking policies via a no-code/low-code solution, leveraging an optional AI model trained with privacy expertise. Automatically deploy them enterprise-wide with broad support for data sources including high-volume repositories.

Ensure Referential Integrity Across Your Data

Maintain data consistency and utility with masking that preserves relationships across data sources, data pipelines, and clouds. Delphix ensures that a masked value remains the same everywhere it appears, protecting the logical integrity of your data so development, testing, and analytics can proceed with production-like quality.

Integrate Data Masking and Delivery

Streamline your workflows by integrating automated masking directly into your data delivery processes. Delphix lets you provision compliant, high-quality data to DevOps and AI pipelines on demand.

Discover how Delphix data masking lets you secure data and accelerate innovation.

Request a Compliance Demo

www.perforce.com/products/delphix/demo/compliance

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

© Perforce Software, Inc. All trademarks and registered trademarks are the property of their respective owners. (0920NL25)