IMPERVA®

Data Masking Best Practices — Five Steps to Making Data Masking a Reality

WHITEPAPER

Executive Overview

Stories consistently appear in the media reporting the negative impacts of data breaches. Data breaches can be detrimental to an organization and can include loss of customer confidence, poor corporate image, a drop in stock price and long-term repercussions resulting from exposed trade secrets. Additionally, non-compliance with security and privacy regulations can come with a hefty price tag.

Despite heightened awareness of these threats, enterprises continue to fall victim to data breaches. While external threats are often sensationalized, the real threat to sensitive data is from insiders. The threat of insider data theft is a very real one with employees stealing trade secrets, business information such as price lists, customer information, source code or business plans. The harsh reality is that in most cases users had authorized access to the data they stole. Data masking offers a way to ensure fewer individuals can access sensitive information within the organization and protect it from falling into the wrong hands.

Introduction: Data Masking

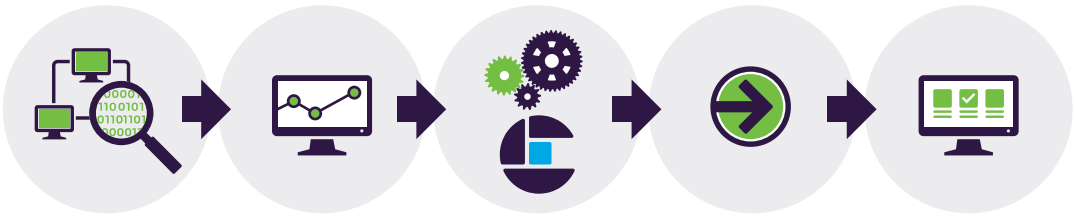
An increasing number of enterprises are relying on data masking to proactively secure their data, improve data security compliance mandates and avoid costs associated with data breaches. Data masking protects data by de-identifying sensitive information contained in non-production environments and enables enterprises to extend their traditional security platforms using a proven technology.

When compared to homegrown data security techniques, data masking represents a paradigm shift in how sensitive data is secured. Masked data retains the statistical properties, integrity and realism of original data, allowing for effective and efficient testing, development, training, and analytics and eliminating the risk of disclosure of sensitive data.

Imperva Camouflage offers an industry recognized best practice methodology that includes five core steps and a logical flow from the discovery of all sensitive information - through to the configuration of the masking engine with appropriate data masking techniques to protect that sensitive information.

The Data Masking Best Practice Includes Five Steps:

1. **DISCOVER:** Identifies data that must be masked without compromising data utility.
2. **ASSESS & CLASSIFY:** Establishes the criteria that will be used to mask the data and create context around the information found in the Discover step.
3. **CONFIGURE:** Creates data masking configurations based upon customer-specific masking requirements.
4. **DEPLOY:** Puts data masking into action with a plan for integrating data masking into the production-to-nonproduction data transition process.
5. **MAINTAIN:** Ensures long-term masking success with knowledge transfer, refresh procedure documentation and change management procedures.



Step One: Discover

Where To Look and What To Look For

To get started with data masking, the first step is to identify the data that must be masked to sufficiently protect the data without compromising data utility.

To mask data properly, one must understand how an organization manages and utilizes its non-production environments. Requirements will vary greatly based on the size of the organization, complexity of data, location of data (i.e. on premise, cloud), scope of data masking and more.

Key activities that will take place during the Discover phase include:

Understand the Drivers for Data Masking, Key Stakeholders and Sponsors

Understanding the organization need for data masking and involving the key stakeholders and sponsors at the onset of an implementation are critical to the success of any data masking project.

Determine Which Data Sources will be in Scope for Masking

A key part of the process is to define the scope of the project at the data and application levels. This should also consider imported and exported data sources ('upstream', 'downstream'), potential referential integrity requirements across platforms, and the various requirements for non-production versions of production data.

Define the Categories of Sensitive Data that Need to be Located

Referencing the documented business requirements and adhering to legislation policies applicable within the organization will assist with defining the sensitive data that needs to be located. Data classification plays an important role in determining the sensitivity of various data types, its intended use(s) and how it will ultimately be masked.

An Examination and Understanding of the Organization's Policies and Procedures as it Relates to:

- **Data and database security:** Security and compliance policies and procedures in place within the organization must be reviewed. Areas such as how database copies are handled, how databases are created, who manages or handles this information, the number of copies and how often they need to be refreshed must be understood. This will help identify any possible gaps that may exist in the data masking process.
- **Governance, risk and compliance (GRC) standards:** Governance, risk, and compliance requirements are critical to how data is used and managed within an organization. GRC typically includes activities like corporate governance, compliance with government standards and regulations, and enterprise risk management. Data masking project requires that you review these requirements carefully.
- **Service level agreements regarding the delivery of masked data** (as this could impact the hardware and infrastructure needed to provision masked data): Typically, performance is identified as one of the key performance indicators (KPIs). Therefore, service level agreements must be reviewed and taken into consideration from the very beginning to ensure the infrastructure is in place to meet the service level agreement.

Step Two: Assess & Classify

What to Mask and How to Mask It

The second step puts in place the criteria that will be used to mask the data and create context around the information from the Discover step.

Key activities that will take place during the Assess & Classify phase include:

Validate and Categorize the Sensitive Information

Some types of data are more sensitive or valuable than others, and data classification hierarchy reflects those differences. An appropriately applied data classification hierarchy can help ensure that only data that actually needs to have rigorous controls is subject to those measures. Some examples of how sensitive data may be classified include the following:

- By policy – e.g. Public, Sensitive, Private, Confidential.
- By compliance regulation or standard – e.g. PCI, HIPAA, SOX, GDPR, etc.
- By business unit or line of business – e.g. commercial markets, consumer markets, etc.

Data classification plays an important role in determining the sensitivity of various data, its intended use(s), and how it will ultimately be masked.

Establish Acceptable Risk Profile

The objective is to establish a risk profile based on the company's risk tolerance. With this risk profile, the organization will decide what constitutes an acceptable level of masking for their environment and select which rows and columns to mask.

Select Masking Strategies for Each Category of Sensitive Information

Having identified the various fields to be masked, the emphasis shifts from 'what to mask' to 'how to mask it'. Any number of data masking transformation techniques can be applied to this process, but an understanding of the implications of the strategies is imperative.

Determine Detailed Masking Requirements, Including Consistency

The purpose of defining functional data requirements is to ensure that all data relationships are factored into the masking process and that the application will work after masking. Failing to do so can have serious impacts on application testing. The major types of relational integrity to be considered are database, application and enterprise level consistency.

Analyze Business Process Impacts, Begin Change Management Planning

A full examination should be undertaken of the infrastructure that will be available during Step 3 below that focuses on creating data masking configurations. At this point, you need to fully understand the uses of the post-masked data for the various environments within the enterprise. Questions should be answered in the following areas: the criteria for users, numbers of copies of data, access, issues of performance and capacity planning.

Step 3: Configure

Configure and Refine Masking Rules

The goal of the Configure step is to create data masking configurations based upon customer specific functional masking requirements defined in prior steps.

During this step, data masking configurations will be integrated into the overall refresh process for non-production environments. This step also provides an opportunity to develop data masking schedules and establish appropriate change management processes.

Key activities that will take place during the Configure phase include:

Configure Masking Rules

This activity is designed to configure the data masking configurations that will be integrated into the overall production to non-production data transition process.

The initial masking configurations define table and field level details, database and application relationships, and provide the basis for creating masking rules. Once these decisions are made, masking algorithms and methodologies are selected that best fit the environment.

Review Performance Considerations

It is important to review performance considerations during data masking configuration and adjust performance settings accordingly. The performance settings can be adjusted at multiple levels of the application stack including the database server, the Imperva Camouflage application server and within the masking engine to maximize performance.

Finalize Solution Architecture

Data masking is incorporated into the overall process for updating non-production environments. In this step, you will define schedules for running data masking and identifying any changes in policies and procedures due to data masking being introduced. As an example, service level agreements related to the time required for the masking process to run may need to be documented.

Finalize Staging Environment

Review the resource allocation (CPU/RAM/storage) for the database staging server as it should be sized according to the volume of the source data and configured to support the bulk nature of the data masking process.

Perform QA & Acceptance Testing

Testing and QA are required to ensure the masking configurations chosen meet the functional masking need and produce the correct results. The scope of acceptance testing and the appropriate sign-off criteria should be agreed upon during the Discover phase.

Update Business Processes Accordingly

Based upon the results of the testing and QA, business process may need to be changed. The documentation related to all masking strategies of sensitive data identified should be updated accordingly, as well as any changes within the operational guidelines.

Step 4: Deploy

Automating and Integrating

With all the planning done, the Deploy step is about moving your data masking into the overall production-to-non-production business process. During this step, data masking schedules are determined and integrated into the existing change management process.

Key activities during the Deploy step include:

Define Masking Schedules

Job Scheduling Scripts are created to facilitate the masking process by automating it for the purposes of creating development, testing QA and training environments unattended. The job scheduling process calls the data masking software to run the appropriate script during the creation of the masked database.

If masking multiple data sources, determining a proper execution schedule will allow you to maximize hardware by running specific tasks in parallel.

Updates to Data Masking Configurations

A data masking project is a set of managed assets and code. As an organization's database structure changes, so will data masking configurations if new personally identifiable information (PII) is introduced. Appropriate policies and procedures must be implemented to facilitate updates to the data masking project assets.

Step 5: Maintain

Ensure Long Term Masking Success

The goal of the Maintain step is to ensure long-term masking success. Knowledge transfer, refresh procedure documentation, change management procedures, and hand-off are key to successful completion of the rollout and to long-term masking success.

Appropriate Knowledge Transfer

Prior to handoff, in-depth training and knowledge transfer should be conducted with development and testing teams. It is also important that post implementation training and knowledge transfer be conducted for new resources to ensure they have an understanding of the data masking process.

Refresh Procedure Documents

All data masking changes/additions identified during a refresh must be documented.

Change Management Procedures and Hand Off

Once masking configurations have been validated for appropriate coverage and environmental fit, they are deployed for use (re-use) in non-production environments. Change management and other maintenance activities are performed on an ongoing basis to identify changing requirements as applications evolve to guide appropriate modification to masking rules.

Summary

Amid growing compliance legislation and an ever-increasing call from consumers to protect sensitive data, organizations need solutions that enable them to protect information without sacrificing the productivity of their employees. Data masking goes beyond access control to protect data from internal users by masking copies of the most current data used in production databases. It enables the creation of fully functional and realistic data. Once masked, the data retains its representation without disclosing the original 'real' information.