IMPERVA®

A BEGINNER'S GUIDE TO DATA MASKING

GETTING STARTED WITH PROTECTING SENSITIVE DATA

WHITEPAPER

Introduction

While most data breaches reported center around malicious external attacks, the true threat to today's organizations may in fact be from within. These insider threats should not be underestimated. A recent study from the Open Security Foundation found that while insiders accounted for 19% of incidents, they were responsible for 49% of the data exposed.[1]

2015 was a record-breaking year when it came to the number of reported data loss incidents. There were 3,930 reported incidents, exposing 736 million records.[2] Despite increased vigilance by organizations, data loss is a very real threat to a company's reputation, corporate image and bottom line. The financial repercussions of this type of data loss are extremely high with costs to the organization estimated at an average total cost of US $4 million, plus possible jail time.[3]

Current approaches to data protection such as firewalls, encryption and passwords fail to sufficiently lock down data. Enterprises today need to go beyond traditional security measures. They require a comprehensive data security solution that includes proactive security monitoring at the data level along with data masking. Together, database audit and protection (DAP) and data masking offer a solution that can safeguard data from theft while maintaining operational efficiency.

Data masking fills the void left by traditional approaches to data protection, enabling organizations to proactively secure corporate data, improve data security compliance and lower costs associated with data breaches. This paper will explore how data masking can help organizations defend themselves against data breaches by de-identifying sensitive information contained in non-production environments.

Data Loss in 2015

3,930
Reported Incidents

736,000,000
Records

A Record Breaking Year[2]

[1] Report "Data Breach Quick View", Open Security Foundation, 2015

[2] Report "Data Breach Quick View", Open Security Foundation, 2015

[3] Report "2016 Cost of Data Breach Study: Global Analysis", Ponemon Institute, June 2016

What is Data Masking?

Data masking, also referred to as data de-identification, anonymization, or obfuscation, is the process of obscuring sensitive data to prevent it from being exposed to individuals not authorized to view it. These individuals are either internal employees, like application developers and testers, or can also be outside consultants or offshore firms.

Gartner defines data masking as "a technology aimed at preventing the abuse of sensitive data by giving users fictitious (yet realistic) data instead of real sensitive data."[4] Properly masked confidential data is safe for use with in-house or third party application development, testing and training.

Sensitive data such as Personally Identifiable Information (PII), Protected Health Information (PHI), Payment Card Industry (PCI) Cardholder Data, Corporate Financial Information (CFI), or Intellectual Property (IP) is identified and moved from or obscured within a database. This may include, but is not limited to: employee records, student records, customer details, electronic medical records, merger and acquisition information, patents and trade secrets, and any other information that may be critical to the health and stability of the organization and protection of constituent's privacy.

With data masking, transformation algorithms are applied to produce fictional but contextually accurate data that is substituted for the original source data. The data looks real, but users of a database would never know that the data has been masked. Masked data provides the most effective way to protect privacy and support compliance initiatives for copies of production databases used for application development, testing and training.

Balancing the need to secure with the need to use.

Data security is a careful balancing act for every organization. While data security is a top priority, it cannot be implemented in such a way that it impacts employee productivity or data realism. For security to be truly effective, it must balance the need to secure with the need for data utility.

Data masking offers organizations a way to carefully balance these competing needs so that data can be used without creating vulnerabilities. The de-identification of data makes data usable without leaving it exposed.

According to Gartner, "masked data must not break application integrity. That is, it should satisfy the same business rules as real data (e.g. masked age is still in the same age group; zip code has the same geographic dispersion; checksum for credit card calculates correctly). This is to ensure that the application running against masked data performs as if the masked data is real and to ensure there are no limitations on a user's abilities to adequately use applications."[5]

Gartner's point identifies a significant challenge for enterprises when it comes to data security. For it to truly be effective, it must protect the data while providing access to realistic data so that application developers or data analysts can effectively perform their work. The concept of data utility is critical to effective data security, which is why a growing number of organizations are including data masking as an essential part of their broader data security strategy. It offers organizations an effective way to protect sensitive data in a test database in advance–even before testing starts.

[4] Report "Magic Quadrant for Data Masking Technology, Worldwide", Gartner 22 December 2015

[5] Report "Static and Dynamic Data Masking Explained", Gartner 20 October 2015

WHY DATA MASKING?

Most organizations use multiple copies of production data to support development and testing in non-production environments. Gone are the days when freely copying production data for use in development and testing was 'the norm'. That throwback to data management is not only extremely risky, it is now even prohibited and penalized by industry regulations such as SOX[6] and HIPAA[7].

Industry experts agree that data masking is a necessary part of an enterprise's data security. The Gartner Market Guide for Data-Centric Audit and Protection report, released in December 2015, includes data masking as a key capability of a broader data security governance strategy.[8]

Data masking enables organizations to eliminate sensitive production data from testing and development environments. The use of production data in non-production environments violates data privacy laws and regulations. Traditional security measures used to protect production data simply do not work for development and testing because testers and developers need access to usable realistic data to perform their jobs. But using real data could result in a privacy violation or data breach. Furthermore, developers and testers require access to valid data to accurately test and deploy their applications.

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, thus allowing effective and efficient testing and development, and research, while eliminating the risk of disclosure of sensitive data. The time consuming and resource intensive nature of home grown data protection solutions ultimately led to a new set of costly problems associated with repeatability and accuracy.

The Dollars and Cents of Protecting Consumer Privacy

- Protecting sensitive data needs to be a strategic imperative for every enterprise and with good reason when you consider the costs of leaving consumer data exposed.
- The global average cost of a data breach has an average percapita cost of $154USD. However, the US and Germany experience higher than average costs of at $221 and $213, respectively.*
- The notification costs of a breach are significant. Ponemon Insititue found that US. organizations spend $2.3M USD on average on notification and post-breach clean up.*
- A single breach is enough for consumers to lose confidence and stop doing business with an organization. For US organizations, the cost lost business after a data breach is $3.97M.*

* Source: Ponemon Institute: Cost of a Data Breach Study, 2016

[6] The Sarbanes Oxley Act 2002, available at www.soxlaw.com

[7] Summary of the HIPPA Privacy Rule, US Department of Health and Human Services

[8] Report, "Market Guide for Data-Centric Audit and Protection", Gartner, 15 December 2015

Considerations for Masking Data

Before getting started with data masking, several factors must be considered.

First and foremost, organizations need to consider the impact on development in terms of the different components of a database application, different testing cycles, types of resources involved, and the overarching business process and technical environment that the software applications reside in.

Data masking may require a shift in thinking for those involved in such organizational activities as development, unit testing, user acceptance testing, integration/system testing, quality assurance, training, parallel testing, production support, and conversion/upgrades implementation.

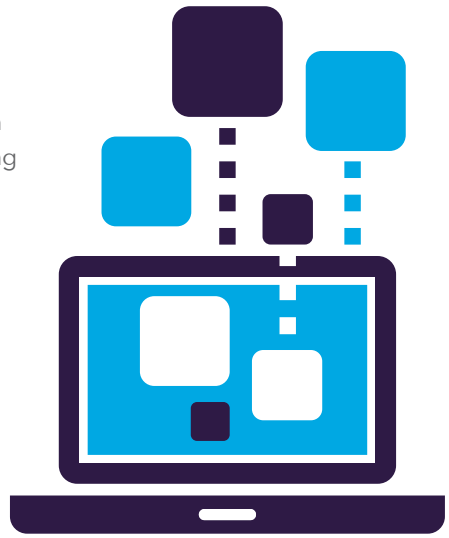
Testing Impacts

Is current testing based on identifiable "individuals"? Often, testers will perform testing using pre-identified individuals with known data characteristics.

Depending on how the data is masked, these "individuals" who were previously identified in production data will no longer be identifiable because data masking will remove or alter identifiable characteristics. Hence, the masked data will no longer have the same "individuals" to test against. In such a case, testing must be approached from a more generic point of view where the focus is on the class (or category) of test cases instead of the individual such as approximate geographic locations, gender, or telephone area code–whatever non-confidential attributes define the scenario under test. This means that other non-confidential descriptors may need to be used to identify the appropriate test cases.

Higher-level considerations when determining the potential impacts of using data masking include:

- What specific types of data actually need masking?
- Is development and testing integrated across multiple databases?
- What is the downstream impact of data masking on development and testing?



Application Impacts

When determining the detailed impacts that data masking might make on various application components, it may be helpful to consider the following elements that shape the technical application environment(s):

- User Interface
- Business Logic
- Reporting
- Queries / Analysis
- Batch Processing
- Interfaces
- Workflow

User Interface

For the development and testing of software applications, a user interface typically requires realistic looking data but not real data. However, that is not always the case.

Business Logic

Is the 'glue' that holds the applications together dependent upon certain data patterns/ characteristics? There are often data edits that must be accounted for when performing data masking to ensure business logic functions correctly.

Reporting

Is there hard coding in the reports? Although not desirable, it is sometimes unavoidable. If this is the case, some analysis may be required to determine if the hard coding is significant and if the data masking configuration must account for it.

Queries / Analysis

Similar to reporting, considerations must be made for ad-hoc querying/reporting. What are the typical parameters used when creating queries? Are there certain data categories that must be preserved?

Batch Processing

This is another significant area of application functionality that requires testing, as significant data volumes are often involved. Are there billing statements to be generated or payrolls to be calculated? How are these calculations/batch processes verified? How will they be verified once the data is masked?

Interfaces

Inbound/outbound data feeds must also be tested. Consideration must be given to how third parties who receive the data feeds handle masked data.

Workflow

In a technical sense, workflow routings/entities may need to be masked for testing purposes. As one example, if application events trigger email messages to clients, the email addresses will need to be valid for testing but modified to prevent test messages from being inadvertently sent to real users or clients.

Beyond the Application

Outside the application itself, several other items must be considered when masking data:

Business Processes

What types of data are collected and is there any sensitive data that employees or contractors have access to that they do not need to access to perform their jobs?

System Integration

Once the data masking implementation has been configured, consideration must be given to how it should be integrated into the database refresh cycle. This typically involves using the command line interface (CLI) to mask database(s) post-refresh.

Security / Authentication

How is database access controlled? Should application profiles be masked–do they contain sensitive data?

Performance (masking)

How time sensitive is the development/testing cycle to refresh times? Data masking software can be sized to help improve masking performance but the processing window(s) should also be clearly established to allow sufficient time for masking.

Size Constraints

Some data masking solutions require a full copy of the underlying production database because they mask 'in-place'. Typical development and testing environments already have sufficient space for one or more such copies, as production data is often used for these activities prior to the implementation of data masking.

Evolution /Growth of Databases and Applications

As applications change, so do data masking requirements. New (sensitive) data fields may be added to applications and data structures may be created, modified or removed.

Benefits of Data masking

Data masking is currently being used by organizations around the world to ensure data is secure. It offers many benefits including:

- Removal of sensitive data from development/testing/training environments, enhancing data security for application development processes.
- Usable, realistic data for application developers/trainers/ testers.
- Assurances that masking at the application level complements masking at an enterprise level.
- Assistance to organizations in meeting regulatory compliance requirements.
- Protection against damaging attacks in non-production environments.

Data masking also protects specific business operations from exposure to sensitive data and optimizes IT staff efficiency. The confidentiality, integrity and availability of information and services are preserved, and the productivity of IT staff in development and testing activities is maintained, along with security administration efficiencies.

Data masking acts as a deterrent to insider threats from privileged or non-privileged users as they simply cannot access sensitive data such as customer information, trade secrets or competitive sales information.

Summary

Data masking offers organizations of all sizes a highly effective way to address data security requirements. The cost of dealing with data breaches is simply too great for any organization to ignore.

While it may require a change in mindset in how data is secured, data masking can swiftly reduce the risk of data breaches. As customer data and other sensitive information is de-identified, a large number of malicious and accidental threats are simply eliminated, helping to protect organizations from the very real threats of insider data leakage or theft.

Next Steps: Learn More About Data Masking

Through its work with organizations around the globe, Imperva has developed proven best practices to help enterprises successfully implement data masking. An established market leader in data masking, Imperva works with enterprises across all industries to deliver cost effective and highly secure data masking solutions.

Data masking best practices along with comprehensive enterprise-grade data masking technologies from Imperva enable organizations to protect against data breaches from the inside. Some of the world's most security conscious organizations currently rely on deep domain expertise and capabilities from Imperva.

Learn more about how data and application security solutions from Imperva help ensure information security and regulatory compliance. Visit us at www.imperva.com, or contact an Imperva sales representative at sales@imperva.com or call 1.866.926.4678.