opteamix

Data Masking Strategy

CONTENTS

Overview

It is shocking to note that about **3.5 billion people** saw their personal data stolen in the top two of the 15 biggest breaches of this century alone. With the average cost of a data breach exceeding $8 million, it is no wonder that safeguarding confidential business and customer information has become more important than ever. Furthermore, with stricter laws and governance requirements, data security is now everyone's responsibility across the entire enterprise.

However, that is easier said than done, and for that reason, an increasing number of organizations are relying heavily on **data masking** to proactively protect their data, avoid the cost of security breaches, and ensure compliance.

This whitepaper aims to give an overview of the world of data masking by uncovering the types, techniques, strategies, and use cases of custom implementations that will help you understand its enterprise potential.

As the name implies, data masking (also known as data obfuscation, data anonymization, data munging, data pseudonymization, data scrambling, data scrubbing, etc.), is a process that organizations use to secure their data. The real data is obscured by random characters or other data that covers classified data points from those who do not have permission to view it.

Why **Data Masking?**

The primary function of masking data is to protect sensitive, private information in situations where it might be visible to someone without clearance to the information. Organizations copy millions of sensitive data to non-production environments and very few succeed in protecting this data when sharing with external and third-party systems too.

Let us consider the following instances:

- **Health care organizations** share patient data with medical systems or healthcare providers for enhanced availability, the integrity of data on electronic Health Records (eHR), research on clinical trials, efficient medical treatments, etc.

- **Financial services institutions** such as banking, mortgage, insurance, etc. depend on rapid transaction speed and global interconnectivity. Unfortunately, all of their critical support infrastructures like trading platforms, central security depositories, payment and settlement systems, and central counterparties each serve as a single point of failure. Therefore, a security breach at any one of those infrastructures could have far-reaching consequences.

- **Enterprises** share data from their production applications with other systems for various business needs, data & business analytics, or to allow a system administrator to test, apply patches, fixes, and run upgrades.

- **Application developers** need an environment-mimicking production setting to build and test new features. This helps businesses achieve a competitive advantage and stay up to date.

As you can see, data masking is essential in many regulated industries where business-confidential information must be protected from overexposure. With data masking, sensitive data categories that include protection of intellectual property, personally identifiable information, protected health information, financial information, and payment card information can only be viewed by the people who are authorized to see it and can see only what they should. By masking data, the organization can expose the data as needed to test teams or database administrators without compromising the data or getting out of compliance. **The primary benefit of data masking is reduced security risk.**

Government Laws & Regulations

The need to protect data and mitigate the risk of compromising critical and confidential information is at the forefront of governments and enterprises, which is why laws, regulations, and business policies have been built around processes. Data masking is the first step to being in compliance with privacy laws, discretion, and fulfilling the obligations of confidentiality.

Some of the regulatory efforts that have been put forth are - The Health Insurance Portability and Accountability Act (HIPAA) which alludes to the protection and confidential handling of protected health information, the Sarbanes-Oxley Act (or SOX Act) aims to protect investors by making corporate disclosures more reliable and accurate and the Payment Card Industry (PCI) Data Security Standard (DSS) which is enforced by Visa and Master Card ensures cardholder data security and adopts data security measures globally.

The **Challenges**

Organizations have started taking threats to data breaches very seriously and have set out to address these issues as quickly as possible. The typical approaches to data protection such as firewalls, encryption, and passwords fail to sufficiently lock down data. Enterprises today must go beyond traditional security measures and instead opt for a comprehensive data security solution that includes proactive security monitoring at the data level.

The idea of merely removing sensitive information from the non-production environment seems to be a simple ask, However, it can pose serious challenges in various aspects. The immediate challenge lies in finding the perfect balance between conformance with privacy laws, discretion, and the obligations of confidentiality. Ensuring data integrity between support and the production environment delivers more value to the business and speeds up releases.

Some of the immediate challenges are:

- Applications are getting more complex and integrated. Knowing where the sensitive information resides and what applications are referencing this information becomes difficult. Because of the ever-evolving nature of applications, maintaining meta-data knowledge of the application architecture throughout its lifecycle also is of concern.

- Since application design is usually not built with the intent of 'hiding' sensitive data from specific environments, the analysis of the dependencies on each of the sensitive elements (such as development and testing needs, target user community, location and compliance mandates to name a few), has to be considered.

- Maintaining application integrity, auditing needs, acceptable performance, and reliability while formulating a flexible solution.

- Repeated masking and synchronizing poses a challenge when auditing is concerned. Keeping track of what has been changed becomes an important business control requirement to prove compliance with regulations and laws. To implement these types of controls, reporting becomes paramount, along with the separation of duties and role-based permissions.

- Complexity in introducing data security as a part of the development life cycle as data models and data stores have evolved with time.

- Providing a flexible solution that can evolve with the application and can be extended beyond the database to logs, XML files, JSON data within an enterprise becomes an important challenge to address.

- Inherent complexities due to referential nature of data, availability of data in multiple tables in de-normalized databases, distributed data sources.

- Diversity in environments with a single point of ownership for each application.

Our **Solution**

Our solution to these challenges regardless of the industry is to apply **our Masking Approach** that is carried out in the following stages:

Discover: Do a thorough discovery of understanding the business, go through the data model, relationship, usage, etc to identify the sensitive data.

Assess: : Define the type, the technique, and where to apply the masking to be adopted for the use case and derive the masking definition for each of the data categories.

Apply: Apply the derived masking definition to the data source.

Validate: Utilize the application test suite/automation testing framework to validate the system behavior and security of information.

***Roles Involved**: This process will involve the expertise and coordinated effort of DB Admin, Security Admin, Application Development, and Testing team.*

Parameters for Solution Identification

Business rule	Analyze the underlying business rules of the use case of information.
Referential integrity	Must maintain the data integrity across tables, databases, and systems.

Row-Internal synchronization requirement	Similar to referential integrity, however integrity within the row or tuple. Ex. PersonName is the concatenation of FirstName, MiddleName, LastName.
Application type	COTS or Custom-built: Analyse if there is any ready solution in the market for business case.
Cost	Must be a cost-effective, yet efficient solution.
Turn-around time	Ready solution masking might be generic and custom solution can target for a specific case, hence might support reduced cycle time.

Where to Apply **Data Masking**

Based on where we apply masking, there are two major types:
Static and Dynamic data masking

Static data masking - This kind is applied to the copy of a single source of data, however, does not limit to the DB but also evolves with log files, XML, JSON, payload files, etc. Admins, typically load data backup into an integrated environment, carry out filtering on the dataset to limit the necessary data needed for testing or business operation, and apply data masking at static and then push to the intended environment.

Dynamic data masking - Dynamic Data Masking, applied on one record at a time, at runtime, dynamically, and on-demand. This process is crucial in enabling continuous deployment for huge integrated applications by avoiding the time spent on creating backup and load to a special copy of the database.

Systems where there are feeds from production at a constant pace to development making it complex to remain compliant, on-the-fly data masking becomes essential in such cases. Further, on-the-Fly data masking happens in the process of transferring data from environment to environment without data touching the disk on its way.

How to Apply **Data Masking**

Data Masking depends on your business needs and rules, as well as applicable data privacy law(s). At a technical level, that usually means deciding how the resulting masked data or, ciphertext needs to appear, if it needs to be reversible or unique, how secure it is, and possibly, what kind of resources and time are available for the process.

Some of the various techniques are:

Data Masking in Action

Opteamix successfully performed the Data Masking Analysis and defined strategies for one of the leading US-based Human and Health Services providers whose business involved crucial and critical rules and regulations for the protection of PII/PHI information. The business needed a solution that would mask sensitive information, retain referential integrity, maintain row internal synchronization, enable continuous deployment of the applications all while reducing time and DBA workload.

A number of their existing applications were data masked as a security-related solution provided by their vendors. However, because of the sheer size of the application, problems arose in the form of high turnaround time and restrictions on certain data records which were either nullified or filtered out even though they were crucial in running the business in onsite-offshore model.

To find a solution to the above-mentioned problem, it was important to recognize the fact that a single solution might not be feasible in this situation. We had to come up with a solution that can leverage people, processes, and technology for easy integration. Our study included-analysis of production issues, mimicking systems for an exact production-like environment of test and dev, covering tests for patch/fixes/upgrades, quick turn-around test, and development environment provisioning, etc

Solution Options

1. Data Masking Solution for COTS like ATS, ERP, and/or CRM products:

With the out-of-the-box pre-defined masking algorithms, masking common sensitive information, such as social security numbers, emails, credit card numbers, etc. reduced significant time in constructing from scratch. COTS applications can be masked with ready market solutions.

Usage of ready market solutions like: DATPROF Data Masking Tool, Delphix, Oracle solution.

2. Data Masking Solution for Inhouse Products & Applications with Free text and Structured JSON/XML

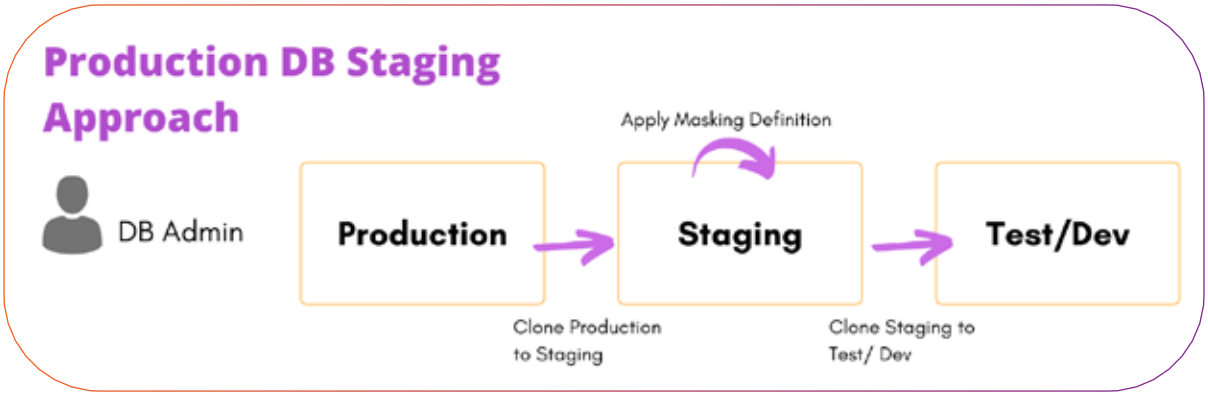
Scenarios where applications have free text of audit notes, logged information, operation team's & application's notes, structured text like Payload, JSON and/or XML, etc.

Due to this complexity, there is a strong need for an Inhouse custom or Hybrid solution after performing the various stages of masking approach of discovery and assessment

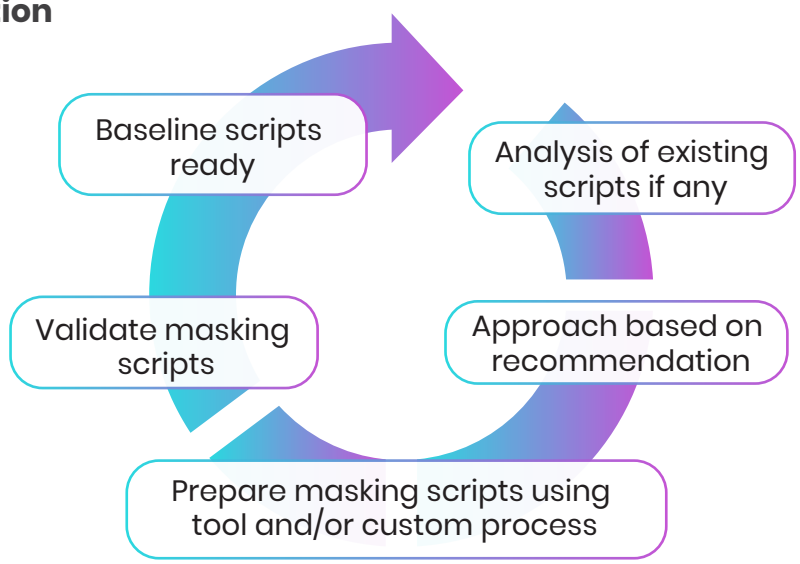
We recommended, Static data masking considering the complexity, cumbersome steps, and tools involved in other types for similar products for one of the Product for supporting US Health Care services.

Identified Techniques to be used are: Substitution, Masking, Numeric/Date Variance and Nulling.

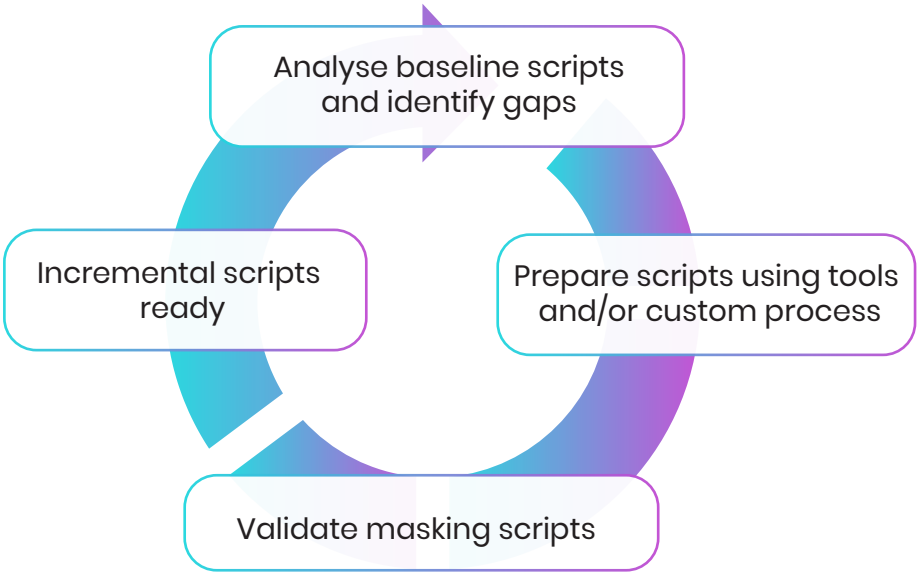
Representation of Incremental approach



Initial Iteration



Subsequent Iteration



Use Case

Take a scenario where application production support is performed by offshore consultants, lending their expertise towards batch job executions, patches, DB upgrades, and data fixes to resolve issues for the operations team. During this process, people who are not authorized to see certain data are privy to lots of information that most often is of a sensitive nature. Due to the ongoing Pandemic and remote working, this process has become more vulnerable to security threats. To mitigate this risk, we carried out the various stages of the masking approach of discovery and assessment and provided the following options to solve the problem:

- **Data Redaction:** Users working on Data fixes are provided with specialized user role and access and rendered with redacted data from DB Client on the execution of SQL queries against the PII/PHI information holding DB tables.
  Example: Email Address of JohnMary123@abc.com can be rendered as JxxxMxxxxx@abc.com or xxxxxxx@xxx.com for the unauthorized user.

- **Use of Datebase Views:** Designed exclusive user roles and permissions for the DB access users and created DB views for the tables which are needed, without the PII/PHI holding columns for Production Support.

- **Application code changes:** Application side code changes for the UI layer to use CSS styles and HTML and JavaScript functions to render masked data for the sensitive information on to the web-application.

Conclusion

With stricter laws and governance requirements, data security is now a concern across industries. Data masking is not domain-specific and offers organizations of all sizes a highly effective method to address data protection. It can be adopted to suit any business needs, enabling organizations to proactively secure corporate data, improve data security compliance, and lower costs associated with data breaches that are 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. Organizations must have a well-defined strategy in place to discover, assess, apply, and validate the right Data Masking process based on business needs.

As customer and organization data and other sensitive information are hidden, 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.

Opteamix is a digital automation technology consulting firm with deep expertise in Application Development, Robotic Process Automation, AI, DevOps, Enterprise Mobility, and Test Automation Services. We are headquartered in Denver, Colorado with a wholly-owned delivery center in Bangalore, India.

Contact Us

USA

9200 East Mineral Avenue
Suite 330,
Centennial
CO 80112

+1 720 508 8780

contact@opteamix.com

INDIA

37/A-07, Southend Road
6th Block, Southend Circle
Basavanagudi
Bengaluru 560004

+91 80 4667 1666

contact@opteamix.com

Visit us at

www.opteamix.com