



Why Add Data Masking To Your Best Practices For Securing Sensitive Data

Dataguise, Inc.

2201 Walnut Ave., Ste. 260

Fremont, CA 94538

(510) 824-1036 www.dataguise.com

Introduction

Databases are critical assets for all organizations. No matter if the entity is big or small, databases are an essential part of any business operation. Most enterprises store sensitive data such as credit card numbers, social security numbers, company confidential data and other data, which make them a target for attacks. A database is like a bank vault that stores money. Databases are critical, data security is a high priority.

Today database attacks, both internal and external, are on the rise. More than 70 percent of all attacks on databases are internal, making them very difficult to detect and curb. Analyst reports indicate that more than 240 million customer records have been compromised to date. Large companies that range from financial institutions, retailers, and insurance companies to prestigious universities regularly suffer negative headlines proclaiming data theft or loss. Such attacks are likely to grow even more in the future unless enterprises take stronger data security measures to protect their databases.

A data security breach can have a severe impact on any organization such as, lawsuits, legal fines, negative brand recognition and decline in stock prices. Mitigating risk is critical to any business. Just like car insurance you may never need it but if you are in an accident you must have it. According to leading industry analyst firms, the cost of a breach is anywhere between US\$50 to \$300 per record, which can collectively add-up to millions of dollars.

The Current Problem: Data is often not protected in non-production environments

Most enterprises secure production databases when dealing with sensitive data, but only a few secure such data in non-production environments such as test, development and Quality & Assurance (QA). Regardless of where the data is stored, production or non-production, data security remains equally important. The value of such data remains the same. Unlike confidential hard copy paper digital information can be duplicated and quadrupled easily, and then rapidly deployed to platforms such as test and development, making the data vulnerable.

To test any business application data is essential. In most cases testing data comes from production environments, making test databases vulnerable to exposure to both internal privileged users and external hackers. Non-production environments such as test, development, QA and staging databases are often given a lower priority when it comes to security.

Multiple copies of production data exists in non-production environments. On average five copies exist for each production database to support test, development, QA, migration and staging environments (See Figure 1). More copies of production data means increased security risk.

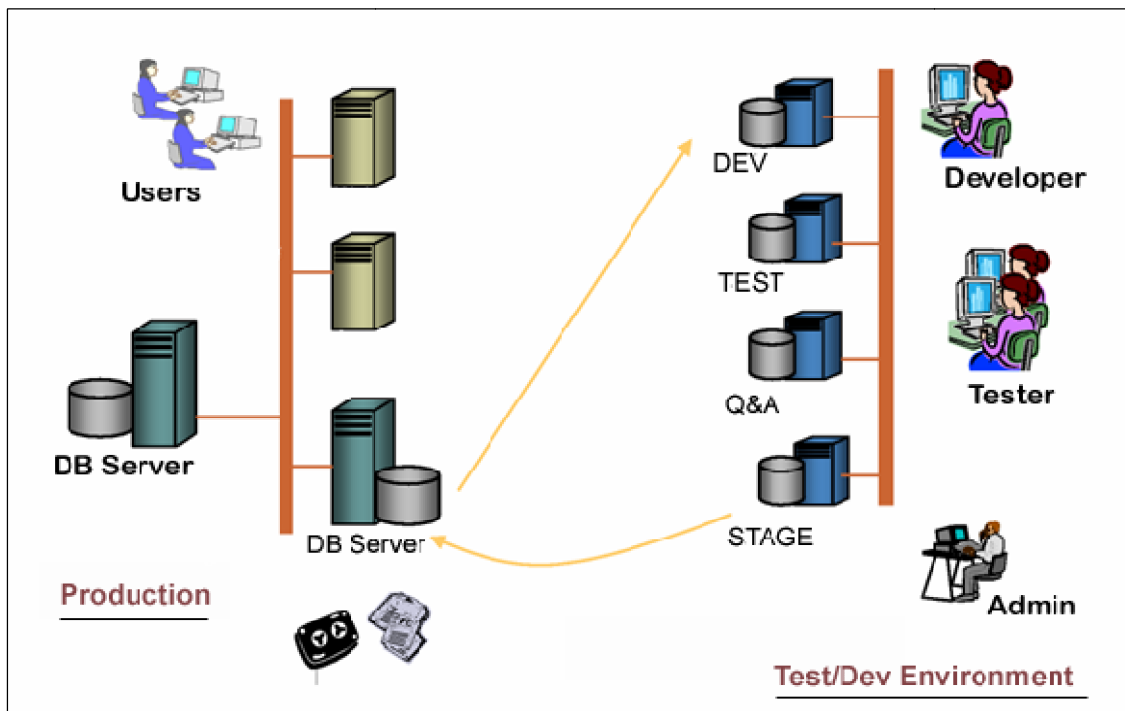


Figure 1. Typical Non-Production Environment

New user accounts created without valid authorization. When a non-production environment is set-up to be accessed by testers, developers, and other IT personnel the new user accounts are typically created with full access to all data. However, the same data in a production environment often has a high level of data access control.

Data is often extracted from non-production databases to other smaller data repositories. Based on requests by customers and other 3rd party entities, many developers and testers extract company data from non-production environments and output them to files and desktops. This practice not only increases data security challenges, but also makes data vulnerable to unintended exposure.

Regulatory Compliance Continues To Pressure Enterprises

Over the past few years, various regulatory compliances such as PCI, HIPAA, GLBA and SOX have put pressure on enterprises. Basic DBMS security such as authentication, authorization, and access control alone are not good enough to meet various compliance requirements. Enterprises have to take advanced security measures such as, data encryption, auditing, data masking, and real-time monitoring to ensure data privacy and protection. However, each compliance requirement is different. Enterprises need to take appropriate advanced security measures to ensure they meet the requirements.

PCI Compliance Mandates Strong Data Security Measures

In 2004, Visa and MasterCard incorporated the Payment Card Industry Data Security Standard (PCI DSS), which is applicable to all major credit-card issuers worldwide. PCI requires that companies establish and maintain adequate internal data security control and procedures pertaining to cardholder information. It covers credit card holder information that is stored or transmitted across the complete technology stack: network, servers, storage, databases, middleware, and applications. PCI concentrates on strong authentication, access control, auditing, and data encryption. It requires establishing strong security policies, processes, and procedures. The top four PCI requirements that require attention from enterprises include:

- **Requirement 3: Protect stored cardholder data.** PCI protecting sensitive data wherever it may be production or non-production, off-line or on-line, on-site or off-site, disk, tapes or devices. Recommended approaches to protect databases include data masking for non-production, data-at-rest and data-in-motion encryption for production environments.
- **Requirement 11: Regularly test security systems and processes.** PCI mandates that all enterprises dealing with credit card numbers regularly test their systems from data privacy point-of-view. Recommended approaches include auditing, monitoring, and encryption.
- **Requirement 8: Assign a unique ID to each person with computer access.** With applications now using web servers and application servers to authenticate users, databases do not have unique IDs to identify users. Therefore, consider integrating application users with database to keep track of who accesses private data. Recommended approaches include auditing and monitoring of sessions across applications and databases.
- **Requirement 10: Track and monitor all access to network resources and cardholder data.** Enterprises need to ensure that only authorized users can access network, this includes monitoring network access and resource utilization. In addition, cardholder data needs to be monitored based on who is accessing or changing such information. Recommended approaches include data masking, encryption and monitoring.

HIPAA Mandates All Patient Records Be Protected In All Environments

HIPAA compliance focuses on protecting patient health information to standardize communication between health care providers and health insurers, and to protect the privacy and security of protected health information (PHI). All PHI-related data residing on any database, backup, tape, or transmitted on network needs complete data protection. The key requirements from a database point of view are in Section 164.308 — administrative safeguards — and Section 164.312 — technical safeguards.

To meet HIPAA compliance requirements, enterprises should first ensure that they establish strong authentication, authorization, and access control security measures, besides having strong policies and procedures. Enterprises should then look at advanced security measures such as data masking and data generation solutions to protect private data in test and development environments. In addition, enterprises should look at data-at-rest and data-in-motion encryption and auditing solutions.

The Solution: Data masking helps protect non-production environments

Data masking, also referred to as de-sensitization, de-identification or data scrubbing, is a process that helps conceal private data. It protects private data in non-production environments such as test, development, and QA, or when data is sent to outsource or offshore vendors. Data masking changes the original data to de-identify it so that it does not relate to any particular person, entity or context. In the data masking process, the data is changed with special characters such as a hash sign or changed with new unassociated data



Figure 2. Data Masking

Data masking defined:

Data masking is the process of concealing sensitive data so that Internal privileged and authorized users cannot access or view the actual data.

The primary focus for protecting sensitive data using a data masking technology is Application and Database Integrity. Sensitive data includes social security numbers, bank account numbers, health related personal information, financial information or any company confidential information. Data masking scrambles data to create new, legible data but retains the data properties, such as its width, type, and format. Common data masking algorithms include random, substring, concatenation, date aging, sequential, and XOR (bit masking).

Recently the adoption of data masking technology has grown mainly because there has been a greater need to protect private data in test environments especially when supporting off shoring or outsourcing of application development. In addition, regulatory and legal requirements are demanding protection for private data regardless of where it is stored. Forrester estimates that 35% of enterprises will be implementing data masking by 2010, with financial services, healthcare, and government sectors leading the adoption.

How Does Data Masking Add to Other Data Security Technologies in Parallel?

Data Masking is different from other data security measures such as encryption, auditing, access control, and vulnerability assessment and monitoring. Each of these technologies play an important part in securing data in production environments, but when it comes to non-production environments data masking alone offers strong protection of private data in such environments.

Data Masking and Auditing

Auditing is a technology that is used to keep track of accuracy of data, such as financial books or company confidential data. If someone accesses or changes data, auditing technology logs that information and provides evidence of the event. Data masking does not keep track of the accuracy of data, or log access information, but focuses primarily on de-sensitizing private data in non-production environments.

Data Masking and Access Control

Access control focuses on ensuring that only authorized personnel can view or change sensitive data. While access control can protect sensitive data in production environments, in non-production such control is not possible. Developers and testers need full access to data, which is where data masking plays a key role.

Data Masking and Encryption

Although there are some similarities between data masking and encryption, they are different in usage, technology and deployment strategies (See Table 1). Encryption can be of two types – data-in-motion encryption and data-at-rest encryption. Both can conceal private data and decrypt it based on encryption keys. Data masking on the other hand conceals private data but cannot de-mask it. Encryption focuses on protecting data from external attacks and breaches, whereas data masking is for protecting against internal users including privileged users. In data masking there is no key management since the focus is not to reverse the masked data to its original form.

Why Is Data Masking Important For All Enterprises?

Data masking is a viable technology to protect private data, especially in non-production environments such as test, development and QA. The key benefits of data masking include:

- **Helps meet compliance requirements.** PCI and HIPAA mandate that any sensitive data in any database or file be secured and only authorized users should have access to such data. Data masking technology helps protect sensitive data in non-production environments such as test and development, by de-identifying data so that privileged users such as testers and developers cannot view it. Testers and developers do not need to view sensitive data for testing or developing an application, besides such data only represents two to three percent of the overall data.
- **Protecting sensitive data in non-production environments.** Enterprises often make a copy of production data and use that for testing of applications or for QA purposes. Data masking helps to ensure that any personal or private data is *masked* before being used in test environments, concealing the data's original value.
- **Minimizing information risk when outsourcing or off shoring.** When outsourcing any application development project or sharing data for any data processing purpose, enterprises should consider depersonalizing sensitive data. Most enterprises rely on trust when outsourcing or off shoring data to vendors, data masking technology helps conceal private data, thus protecting it from being misused or stolen.

What Data To Mask?

Any structured data can be masked, but the focus should only be on sensitive information, which might only be two or three percent of information stored in a database. The goal is to de-identify sensitive information such as credit card or social security numbers. The following are typically the prime candidates for masking:

- **Personal Information.** Regulatory compliance requires that customer information should not only be protected from hackers but also not be accessed or viewed by *privileged users* including developers, IT administrators, testers, DBAs and other IT personnel. Consumer or employee related information including credit card, social security numbers, and addresses fall under the definition of sensitive data.
- **Financial Data.** All ERP systems contain financial data such as business transactions, profit and loss information, discount information, deal size and revenue information. Such information is not required for developing or testing applications therefore should be masked.
- **Company confidential.** In product, any other company confidential information should be masked including future roadmap plans, employee and executive salary information, and blueprints to technologies Sensitive Data Discovery and Classification Is Critical.

For any successful data masking project, careful planning is very important along with a deep understanding of the data and how it relates to existing processes. Without data classification initiatives data masking projects are doomed to fail. Understanding data requires knowing what data is in which column.

Data Masking Algorithms

There are several data masking algorithms that can be used such as:

- **Fictitious data.** This technique substitutes data with some fictitious value, making the data look real when it is in fact bogus. This technique does not typically affect the application or testing requirements because it retains all of the data properties. For example, the customer name "John Barrow" could be substituted with the name "Jim Carlos." (See Table 3).
- **Date aging.** In this technique a date field is either increased or decreased, based on policies defined for data masking. However, date ranges must be defined within acceptable boundaries so that the application is least affected. An example would be moving the date of birth back by 2,000 days, which would change the date "12-Jan-1978" to "16-Mar-1972."
- **Numeric alternation.** In this technique, you increase or decrease a numeric value based on a percentage. For example, a salary value could be increased by 10 per cent. This approach conceals the real value of the data, but if someone knows even one real value, they could decipher the entire pattern. While this is an easy technique to employ, it can also be easily decoded.
- **Shuffling data.** In this approach, data in a particular column is moved to another column or another row. It's like shuffling a pack of cards, and the sequence is broken. For example, moving an account number to a random row, so that John's account number is different from the original.

Algorithm	Original data	Masked Data	Explanation
Fictitious Data	613-30-3291 (SSNO)	613-30-#### (SSNO)	<i>Last four characters hashed out</i>
Random Data	John Barrow	Jim Arthur	<i>Random data</i>
Date Aging	5/1/2006	3/1/2002	<i>Date decreased by 4 years and 2 months</i>
Numeric Alteration	10201	10401	<i>Numeric increment by 200</i>
Shuffling Data	Jack Mellon	Roger Smith	<i>Name was shuffled</i>

Dataguisse Masking Solution Meets Compliance Requirements

Dataguisse offers DgMasker for databases a highly automated and advanced data masking security solution that helps enterprises meet various compliance requirements such as PCI, HIPAA, GLBA and SOX. Unlike other solutions in the market, DgMasker has been built from ground-up with security and compliance in mind. DgMasker lets application developers and testers test applications against production data *without* being exposed to sensitive data such as credit card number, social security numbers, account numbers and other data. This solution helps meet various compliance and auditor requirements. Key features of DgMasker include:

- Highly automated and easy-to-use data masking solution
- ER diagrams depict graphical representation of schema structures and relationships
- Supports referential integrity
- Policy-based solution to adapt to business requirements
- Database agnostic to support any DBMS
- High performance and scalable to support very large environments
- Advanced masking algorithms
- Command line option allows easy integration with batch programs

Conclusion

All enterprises should ensure strong protection for private data regardless of where it is stored – production or non-production environments, to meet compliance requirements and defend against attacks. The first step in this process is the discovery of sensitive data. The risk of not securing any databases in which sensitive data resides is huge. Loss of this data can have a negative impact on the business, including legal and financial losses. Data masking is a viable solution that is highly recommended for use in non-production environments such as test, development, Q&A and staging.