Abstract

Enterprise production environment data contains sensitive personal information that must be protected from data thefts and frauds. Data Privacy regulations have made it mandatory that private data must be managed in a secure manner and should not be exposed to unauthorized users. At the same time, legitimate business activities like application development and testing require realistic data for successful testing. Information systems also need to hide confidential data from less privileged users for outsourcing needs. Enterprises face a daunting task of balancing out the need of protecting confidential data and at the same time making it available to entities that need them.

A good data masking technology enables protection of sensitive data. Infosys MaskIT is an enterprise class, powerful data privacy solution that helps companies to automate masking of personally identifiable information to prevent its abuse and improve compliance.

For more information, Contact askus@infosys.com
Data Privacy has Serious Consequences

Information is the oxygen of organizations. Enterprises need to store and process private information in their production environment databases. This information can consist of employee, customer, partner and vendor records containing sensitive details like names of individuals, addresses, telephones, emails, social security numbers, credit card information, health insurance details, financial records etc.

Individuals whose data is maintained by the companies expect that their personally identifiable information (PII) will be protected using proper measures. This data is required to be shared inside and outside the organizational periphery for legitimate purposes like business application development, testing, training, statistical analysis, business process outsourcing and market research. Such a shared data is vulnerable to data privacy violations if appropriate measures are not taken.

In case a sensitive data leak materializes and results in exploitation, the ramifications for the organization can be disastrous, often resulting in lawsuits, loss of customer confidence, brand damage, erosion of share price, bad press and loss of revenue or in worst case a possibility of business closure.

Application Test Data: A Major source of Data Thefts

Businesses need applications to function and these applications require maintenance and testing. New applications get developed as companies grow. These activities are managed by internal development teams or external contractors or outsourcing vendors.

Application development and testing activities need realistic data for validations. Usually, copies of production environment databases are created using internally developed scripts and given to development teams. However, this method is risky since real data with sensitive information could fall in wrong hands. This also applies to analysts or trainers who need such a data for their work.

This increases the chances of data getting stolen. For example, while testing an online banking system, application tester can update customer records and as a result can view names, addresses, social security numbers and other private information of individuals. It is possible that this tester can steal this data without anyone being aware.

This risk is further elevated due to the fact that just a single copy of master data is not sufficient for application development. Business applications run in multiple environments and on multiple platforms. To achieve productivity, development teams need separate copies for each platform to perform the work in parallel by separate teams. This means there would be multiple copies of production database instead of just one.

As infrastructure security technologies have matured, it has become quite common to implement well known and standard security measures like access permissions, setting up firewalls, enabling database native security mechanisms, conducting audits and deploying automated monitoring tools along with physical security like access cards to protect production environment.

However, such measures are absent in application development environments. There is a lack of audits or monitoring of application testing data. This can result in master data copies being abused. For example an external contractor could misuse stolen retail customer data or an employee might store such data on a detachable hard disk that gets stolen.

The application development data without adequate security measures can also be stolen by a hacker if an entry inside company infrastructure is achieved.

Complying Privacy Laws

Worldwide, there have been numerous data privacy laws introduced by the governments to make sure that private data is responsibly handled and proper care is taken to avoid leaking of this data. These laws have come into existence as a result of many well published data frauds by the hackers and concerned consumers pressurizing governments to take steps to keep their personal information safe. These legislations aim to improve data privacy protection and compliance to safety requirements.

The Payment Card Industry Data Security Standard (PCI DSS) act makes it mandatory for credit card payment processing companies to maintain data confidentiality while storing, processing and exchanging credit card data. Its section 6.3.4 specifies that production data should not be used for testing or development.
The USA Health Insurance Portability and Accountability Act (HIPAA) mandates maintaining privacy of individually identifiable health data. The Sarbanes Oxley Act (SOX) and Japanese J-SOX specify standards on internal controls for corporate data. The Gramm-Leach-Bliley Act (GLBA), also known as the Financial Services Modernization Act of 1999 requires companies to design, implement and maintain safeguards to protect customer information. The European Union Data Protection Directive 95/46/EC regulates the processing and movement of personal data with a requirement to maintain privacy. Australian Privacy Amendment Act, Canadian Personal Information Protection and Electronic Documents Act, New Zealand Privacy Act, Hong Kong Personal Data Ordinance and United Kingdom Data Protection Act are other similar legislations.

These laws have made it compulsory for companies to protect confidential and sensitive information from falling into the hands of unauthorized users. Violations can result in fines and penalties of thousands of dollars per day and even criminal proceedings against executive management with a possibility of a jail time.

**Fears are preventing Outsourcing Benefits**

Businesses want to outsource application development and testing. They are also interested in outsourcing business functions like finance, human resources, accounting, legal, customer service and technical support.

This allows companies to focus on core competencies, lower the cost of operations, tap remote talent and reduce geopolitical risks by diversification.

Many companies would like to take full advantage of outsourcing opportunities but they are afraid to do so as a result of data privacy concerns. For example, it is possible that an employee of an outsourced organization might be able to view sensitive details though information systems while performing duties which can result in data theft and fraud.

**Data Masking is the Best Alternative**

Protecting privacy is an important ingredient of overall enterprise data security plan. It improves customer trust in the brand. The challenge is to provide an appropriate level of protection and at the same time meet business objectives, thus ensuring that data is made available on a need-to-know basis.

There are two approaches to solve this problem. First option is to create synthetic test data using a software tool. This can be done quickly and works well for simple test cases. However, typical application backend data has complex relationships and consists of complicated scenarios. This makes synthetic data creation a complicated task requiring significant manual efforts and a deeper understanding of application internals and business domain. Besides, this approach addresses only a part of the problem and does not obfuscate data at user interface points of business applications which is an important need for outsourcing activities.

The other alternative is to implement data masking. You can either develop a generalized masking solution or purchase an off-the-shelf solution. Developing an in-house system can take months and will require maintenance efforts and specialist staff. An out-of-box product can be customized and deployed in a much shorter time and at a much lesser cost.

**Data Masking Defined**

Data masking is the technique of modifying sensitive data to create life-like but false values using masking software. Its effectiveness depends on whether you can create source values from the target or not. Data masking retains look and feel of data so that correct testing is possible and also less privileged users can continue to perform their work with rest of the
unmasked data.

Masking is also known by other names like data obfuscation, data de-identification, data depersonalization, data scrubbing or data scrambling.

**Selecting the Right Technology**

It is vital to select a proper masking technology for successful and timely completion of data privacy program that meets its goals. A good solution should be able to perform de-identification with consistency and standardization by playing the role of a centralized masking authority for an entire organization and not for individual departments.

The software must support popular databases, packaged applications and operating systems. Private data does not reside in databases alone. It is scattered throughout organizational computers and stored in different types of files. Hence, files are also an important source of confidential data.

Masking solution should make it easy to discover confidential data, its relationships and its constraints across data sources. This is an important requirement during the initial stages of data privacy initiative.

The software should preserve referential integrity so that de-identified data is usable in testing projects. It should support batch processing of masking operation so that execution can be scheduled at lean times.

Production databases are huge in size and are growing rapidly each year. The solution should be scalable and fast enough to scrub large amount of records within a reasonable time.

Regular expression based rules are necessary to perform selective masking for productivity. For example data can be de-personalized based on the country of a customer or an employee since data elements like social security number, zip code have different format for every country.

It is mandatory that a data element like an employee number always gets masked to same value irrespective of where it comes from. This employee number could reside in database or file or an application. This is known as deterministic masking.

Role based access is critical for security and separation of duties. Project configuration capability is necessary for logical division of work and reuse. Reporting is valuable for auditing purposes and to demonstrate compliance.

A rich set of built-in masking transformations (algorithms) is a must, in order to take care of diverse and complex needs. Typical examples include shuffling (like a pack of cards in deck), random value creation (for example replacing New York with Washington) with an option to limit output within a range, date shift, modify by specified value, encryption/decryption using a specified key, hashing (for example a credit card number 9435-2344-5567-3433 is masked to 9435-2344-5567-XXXX), delta increment or decrement, and custom (plugging in user provided algorithm for highly specialized requirements).

The software should also provide ready-to-use transformations that can be applied to common sensitive fields like social security number, telephone, credit card number and email address for convenience.

**Dynamic Masking: The Key Differentiator**

Most of the masking product vendors provide static masking which scrambles data in production environment before handing over to software testers. However, the key differentiator for a masking solution is the dynamic masking capability, also known as on-the-fly masking.

Dynamic masking applies to data in motion as it moves from production environment to business applications that may or may not reside locally. It involves hiding private data from the less privileged users of application software. For example a call center employee should not be able to look at sensitive financial details of a high net-worth individual and yet be able to carry out the duties. This application software could be a desktop client or a web based system.

Data masking engine should be intelligent enough to perform scrubbing based on location and role of the user. It should also consider country specific formats for sensitive fields since these vary for every country.

In order to support dynamic masking capability, the software should be able to plug into strategic locations like a web server or a message queue so that scrubbing can be performed just before a data record is shown to user. This functionality should necessitate very minimal code changes to the application.

Dynamic masking is critical for the privacy initiative and without it, achieving objectives of data protection are almost
Infosys MaskIT -The Smarter Choice

Infosys MaskIT is an easy-to-use, high performance data masking product to address the privacy protection needs of an entire organization in a centralized manner. It is modeled on requirements of a good masking system.

It can be quickly deployed to automate the de-identification process. It ensures data integrity across applications, databases, tables, and columns. It provides deterministic masking and can perform both static and dynamic masking.

This system supports all major databases like Oracle, SQL Server, DB2, Sybase, IMS DB/DC and popular file formats like fixed record length, delimited, XML, SWIFT, FED and CHIPS. It runs on operating systems like XP, Vista, 2003, Linux and UNIX.

As part of dynamic masking needs, the solution can perform real-time masking of messages from TIBCO, WebSphere, MSMQ and real-time masking of HTML pages on web servers like Apache, WebSphere and IIS.

Its graphical workbench makes it easy to discover and model data elements, identify data relationships and create masking configurations. Configurations are stored in XML format allowing further reuse.

More than 20 out-of-box de-identification techniques ensure you will not need to implement your own transformation logic. Rules can be added for more flexibility. You can deploy MaskIT as a standalone application or integrate into other applications.

Capabilities like incremental masking (obfuscating only newer data), de-scrubbing (to generate original values), filtering (masking only the subset), role based access, schedule batch processing and reporting further improve the usefulness of the solution.

MaskIT is a highly scalable system and can mask over a million database records within 2-3 minutes.

Summary

Organizations are increasingly becoming aware of data privacy needs and challenges. They are in need of a right masking solution.

Data masking must be an integral part of enterprise data security initiative to avoid costly downsides. A good solution is critical for the successful implementation of ever changing masking requirements.

MaskIT, with its key differentiator as dynamic masking is a proven masking product that can be quickly deployed to successfully automate complex and diverse data privacy and protection needs of an enterprise.
Infosys among the world's top 50 most respected companies

Reputation Institute's Global Reputation Pulse 2009 ranked Infosys among the world's top 50 most respected companies.

About Infosys

Infosys provides its clients with business and technology consulting services that deliver measurable business value to help you build tomorrow's enterprise. Through a well-integrated end-to-end range of consulting, technology, engineering and outsourcing, Infosys clients derive the business value they’ve always been requiring from enterprise transformations. Our extensive research teams, including the award-winning SET Labs, ensure Infosys solutions are always cutting-edge and relevant. Our high investment in training – over 10,000 graduates a year pass through our Mysore campus, the world’s largest corporate university – ensures our people stay best-in-class. Infosys (NASDAQ: INFY) also believes in giving back to the communities with scholarships, libraries, schools, and many other fields through the Infosys Foundation.

Global presence

The Americas

Atlanta
Bellevue
Belo Horizonte
Bentonville
Boston
Calgary
Charlotte
Chicago
Detroit
Freemont
Hartford
Houston
Los Angeles
Minneapolis
Monterrey
New York
Philadelphia
Phoenix
Plano
Reston
Toronto

Europe

Amsterdam
Brno
Brussels
Copenhagen
Dublin
Frankfurt
Geneva
Helsinki
Łódź
London
Madrid
Milano
Milton Keynes
Oslo
Paris
Prague
Stockholm
Stuttgart
Swindon
Toulouse
Walldorf
Zurich

Asia Pacific

Brisbane
Dubai
Hangzhou
Hong Kong
Kuala Lumpur
Manila
Melbourne
Moscow
Perth
Shanghai
Sharjah
Singapore
Sydney
Tokyo
Wellington

Africa

Reduit