NIIT technologies

Data Masking Techniques for Insurance

Lalit Mittal

CONTENTS

Introduction

Insurance companies are using technology to manage communication, processes, storage and security. Information Technology (IT) solutions are no longer just a business enabler but an integral part of providing enhanced customer service. IT is crucial to managing information efficiently and effectively.

However, an overlooked risk for insurance companies is the vulnerability of personal and business information used for testing and application development. Many organizations use real data during the test and development phase of new software applications. This includes financial records, transac

tional records, and other Personally Identifiable Information (PII). Further, test environments are less secure because data is exposed to a variety of unauthorized sources, including in-house testing staff, consultants, partners and offshore development personnel.

As insurance companies are highly regulated, the use of real data for testing and development may be putting them at risk of non-compliance with international regulations. Research has also found that customer loyalty towards insurance companies is dependent upon the perception that the institution is taking every measure to protect the customers' personal financial information. Hence insurance organizations must have strict governance over:

- Types of real data used in application testing and development
- Information security precautions and responsibilities
- Use of cloud computing, distributed computing and outsourced services,
- Experience with data breaches involving real consumer data

This whitepaper explores the various data masking techniques and the best practices of protecting sensitive data.

What is Data Masking?

Data Masking is the process of replacing existing sensitive information in test or development databases with information that is realistic but not real. **Data masking techniques will obscure specific data within a database table ensuring data security is maintained. The data masking is applied across applications and environments in order to maintain business integrity.**

Data is de-identified to ensure that sensitive customer information is not leaked outside of authorized environments. It must be provided at non-production environments as new data processing environments gets built. The new environment has to be created using data that is authenticated and genuinely useful. This ensures that new initiatives will work as they have been evolved on data that reflects real-world scenarios.

Data masking is proving to be an effective strategy in reducing the risk of data exposure from both inside and outside an organization, and must be considered a norm for provisioning non-production databases. Effective data masking requires data to be altered in such a way that actual values are re-engineered, while retaining the functional and structural meaning of the data, so that it can be used in a meaningful way without compromising on security.

Some key aspects involved in the masking of data include:

- encryption of data
- ensuring relational integrity
- establishing security policies so as to define the boundaries between the administration and the actual user of the data

Data masking solutions should provide for synchronizing between different data types:

- Row-Internal Synchronization
- Table Internal Synchronization
- Table-to-Table Synchronization

- Table-to-Table Synchronization on Primary Key
- Table-to-Table Synchronization via Third Table
- Synchronizing between Different Data types
- Cross Schema Synchronization
- Cross Database Synchronization
- Cross Server Synchronization
- Cross Platform Server Synchronization

Types of Data Masking

There are a variety of masking routines that are used for different purposes. These masking routines are based on the degree of exposure of data and the amount of control maintained. The three masking routines are listed below:

1. Light Masking on a Bug-Fix or Fire-Fighting Database

In order to be effective, light masking on a Bug-Fix or Fire-Fighting Database needs to have as few changes as possible. The items that can be safely masked in a Bug-Fix database include bank account or credit card numbers. These numbers when used as join keys can be used for protection. In general, any opaque information whichngful to an external organization can be masked in these circumstances.

2. Medium Masking on Internal Development Database

Databases that are used by internal development, testing and training departments and have no visibility outside the organization receive medium level of masking. Items like personally identifiable information in databases or sensitive data like bank account numbers are viable for medium level masking.

3. Thorough Masking on an Outsourced Database

When operational control of the test and development databases are handed over to third-party then thorough masking of the data is required. In such a case only real time information needs to be passed to the remote personnel to perform their functions.

Techniques used for Data Masking

Substitution

This technique consists of randomly replacing the contents of a column of data with information that looks similar but is completely unrelated to the real details Substitution is very effective in terms of preserving the look and feel of the existing data. The downside is that a large store of substitutable information must be available for each column to be substituted.

Shuffling

This technique uses existing data as its own substitution dataset and shuffles the data in such a way that the records in the dataset do not reveal protected details. Shuffling is similar to substitution, except that the substitution data is derived from the column itself. Essentially, the data in a column is randomly moved between rows, until there is no longer any reasonable correlation with the remaining information in the row.

Using analytical means, if the algorithm for the data shuffle can be determined, then the data can easily be de-shuffled and its original meaning can be known. However, shuffle rules are best used on large tables, and they leave the look and feel of the data intact. They are fast, but great care must be taken to use a sophisticated algorithm to randomize the shuffling of the rows.

Number and Data Variance

The number and data variance technique modifies each number or value in a column to some random percentage of its real value. This technique is useful for numeric and date data only. For example, the date field could be converted simply to a time zone definition, thus creating a difference in the meaning of data. It offers the advantage of providing a reasonable disguise for the data, while still keeping the range and distribution of values in the column within existing limits.

Encryption

This is a simple and frequently used technique for statistically altering the data making it look realistic. This technique deforms the data and also makes it longer. In order to be useful again, the data needs to be decrypted, hence revealing its original meaning. This technique offers the option of leaving the data in place and visible to those with the appropriate key while remaining effectively useless to anybody without the key. However, it is one of the least useful techniques for anonymous test databases.

Truncation

This is one of the best techniques, and has the added advantage of making the system more sophisticated and capable. Truncation simply removes the sensitive data and retains the meaningful data structure.

However, from a test database standpoint it is one of the least desirable techniques used for data masking. Deleting columns or replacing the values with null is not a useful data sanitization strategy as the test teams need to work on the data or a realistic approximation of it.

Masking Out

Masking means data anonymization where certain fields are masked with a mask character (say X). This technique does not allow anything to be deduced from the database as the data content is disguised while still preserving the look and feel. It is fast and powerful only if the data is specific and invariable. In other cases it becomes a complex and slow process.

Selective Masking

This masking technique applies masking operations to a sample of data in the table. The sampled rows should be retrieved randomly from the entire contents of the table.

Challenges of Masking Data

Organizations have tried to address the following challenges with various data masking solutions:

1. Minimizing risk: No matter what security measures are taken, there is always a degree of risk involved in handling a large amount of sensitive data. Data breaches can damage a company's reputation, increase liabilities and invite legal suits.

2. Maintaining accountability: Data breaches create negative publicity, harm current and future business, and damage organization's reputation and the client's confidence in it. It is crucial that the organization stays accountable to all stakeholders, customers and employees, and addresses their privacy needs effectively.

3. Compliance with regulatory norms: Confidentiality and privacy norms demand the protection of data against theft. Compliance to all norms is essential to prove the company's commitment to its customers.

Complications of Data Masking

1. **Data Utility:** Masked data should look and act like real data. Data must be fit for:
   - Proper testing and development
   - Application edits
   - Data validations
2. **Data Relationships:** Must be maintained after masking on
   - Database level Referential Integrity (RI)
   - Application level RI
   - Data Integration (Interrelated database RI)
3. **Existing Business Processes:** Must fit in with existing IT and refresh processes
4. **Ease of use:** Must balance ease of use with need to intelligently mask data
   - Usable data that does not release sensitive information
   - Knowledge of specialized IT/privacy topics and algorithmic importance should be pre-configured and built into the masking process
5. **Customizable:** Solution/Process must be capable of being tailored to specific needs of the clients

Benefits of Data Masking

- Increases protection against data breaches. This is achieved through-
  - Defined process, procedure, and a mature system to mask sensitive information
  - Enhanced quality of data privacy, by involving compliance and audit officers to preview data masking and protection policies—even before actual data masking takes place

- Implementation of data masking policies and procedures in an iterative fashion
- Use of verified data masking techniques

- Enhanced development, testing and training quality. This is achieved through-
  - Use of data masking rules and techniques to produce high quality test and provisioning data, thus streamlining the development process
  - Customized data protection policies, specific to customer preferences and business requirements, ranging from very liberal to very restricted
  - Use of application accelerators (as per the technology employed) instead of custom built coding and scripting, to lower overall maintenance costs
- Enables off-site and cross-border software development and data sharing
- Ensures compliance with industry certifications like HIPAA, GLBA etc., privacy legislation and policies
- Provides confidence about security issues to clients
- Leverages information sharing

Data Masking Solution

In situations where it is imperative for an organization to share sensitive data, the Oracle Data Masking Pack provides a comprehensive easy-to-use solution, to share production data with internal and external entities, while preventing sensitive information from being disclosed to unauthorized parties. The solution replaces sensitive data in databases with realistic-looking, scrubbed data based on masking rules and conditions. Insurance companies can now use real data to represent authentic application and database scenarios in their testing processes, without violating privacy policies or laws.

The Oracle Data Masking Pack enables end to end secure automation for provisioning test databases from production in compliance with regulations. The pack reduces risk of breaching sensitive information when copying production data into non-production environments during application development, testing or data analysis.

Oracle Data Masking Pack is also integrated with Oracle Provisioning and Patch Automation Pack in Oracle Enterprise Manager to clone-and-mask via a single workflow. The secure, high performance capabilities of Oracle Data Masking, combined with the end-to-end workflow, ensures that enterprises can provision test systems from production rapidly, instead of taking days or weeks as in the case of separate manual processes.

Implementing Data Masking Solution

With Oracle Data Masking, Oracle has developed a comprehensive four-step approach towards implementing data masking, called **Find, Assess, Secure, and Test (FAST):**

Find: *This phase involves identifying and cataloging sensitive or regulated data across the entire enterprise. Typically carried out by business or security analysts, this exercise will come up with a comprehensive list of sensitive data elements specific to the organization, and discover associated tables and columns across enterprise databases that contain the sensitive data.*

Assess: *In this phase, developers or Database Administrators (DBAs), in conjunction with business or security analysts, identify the masking algorithms that represent the optimal techniques to replace the original sensitive data. Developers can leverage the existing masking library or extend it with their own masking routines.*

Secure: *In this step, the security administrator executes the masking process to secure the sensitive data during masking trials. Once the masking process has been completed and verified, the DBA then hands over the environment to the application testers. This step and the next may be iterative.*

Test: *In the final step, production users execute application processes to test whether the resulting masked data can be turned over to other non-production users. If the masking routines need to be tweaked further, the DBA restores the database to the pre-masked state, fixes the masking algorithms and re-executes the masking process.*

Comprehensive Enterprise-Wide Discovery of Sensitive Data

To begin the process of masking data, the data elements that need to be masked in the application must be identified. The first step that any organization must take is to determine which data is sensitive. Sensitive data is that which is specifically related to certain government regulations and industry standards that govern how it can be used or shared. Thus, the first step is for security administrators to publish what constitutes sensitive data and get agreement from the company's compliance or risk officers.

A typical list of sensitive data elements may include:

Person Name	Bank Account Number
Maiden Name	Card Number (Credit or Debit Card Number)
Business Address	Tax Registration Number or National Tax ID
Business Telephone Number	Person Identification Number
Business Email Address	Welfare Pension Insurance Number
Custom Name	Unemployment Insurance Number
Employee Number	Government Affiliation ID
User Global Identifier	Military Service ID
Party Number or Customer Number	Social Insurance Number
Account Name	Pension ID Number
Mail Stop	Article Number
GPS Location	Civil Identifier Number
Student Exam Hall Ticket Number	Credit Card Number
Club Membership ID	Social Security Number
Library Card Number	Trade Union Membership Number

ISOLATING SENSITIVE DATA FOR MASKING

Data Masking provides several easy-to-use mechanisms for isolating the sensitive data elements.

- **Data Model driven:** Typical enterprise applications, such as E-Business Suite, PeopleSoft and Siebel, have published their application data model as a part of their product documentation or support knowledge base. By leveraging published data models, data masking users can easily associate the relevant tables and columns to mask formats to create the mask definition.

- **Application Masking Templates:** Data Masking supports the concept of application masking templates, which are XML representations of the mask definition. Software vendors or service providers can generate these pre-defined templates and make them available to enterprises to enable them to import these templates rapidly into Data Masking, and thus accelerate the implementation process.

- **Ad-hoc search:** Data Masking has a robust search mechanism that allows users to search the database quickly based on ad hoc search patterns in order to identify tables and columns that represent sources of sensitive data. With all its database management capabilities, including the ability to query sample rows from tables, Data Masking can assist enterprise users in rapidly constructing the mask definition—the prerequisite to masking sensitive data.

Using the combination of schema and data patterns, and augmenting them with published application metadata models, enterprises can develop a comprehensive data privacy catalog that captures the sensitive data elements that exist across enterprise databases.

More than being only a static list, this is a dynamic living catalog managed by security administrators. It needs to be refreshed as business rules and government regulations change, as well as when applications are upgraded and patched, and new data elements containing sensitive data are discovered.

Conclusion

Staying compliant with policy and government regulations while sharing production data with non-production users has become a critical business imperative for insurance companies. Test and development databases will require some form of sanitization in order to render the information content anonymous. Oracle Data Masking is designed and optimized for today's high volume enterprise applications running on Oracle databases. Leveraging the power of Oracle Enterprise Manager to manage all enterprise databases and systems, Oracle Data Masking accelerates sensitive data identification and executes the masking process with a simple easy-to-use web interface that puts the power of masking in the hands of business users and administrators.

Organizations that have implemented Oracle Data Masking to protect sensitive data in test and development environments have realized significant benefits in the following areas:

- **Reducing risk through compliance:** By protecting sensitive information when sharing production data with developers and testers, organizations have been able to ensure that non-production databases have remained compliant with IT security policies while enabling developers to conduct production-class testing.

- **Increasing productivity through automation:** By automating the masking process, organizations have been able to reduce the burden on DBAs who previously had to maintain manually-developed masking scripts.

REFERENCES

- Department of Information Technology, Ministry of Communications and Information Technology, Government of India http://www.mit.gov.in/content/view-it-act-2000
- Articles and white papers about Data Masking and Privacy http://www.DataMasker.com
- Oracle Corporation data masking techniques www.oracle.com
- Financial Data at Risk in Development: A Call for Data Masking Ponemon Institute© Research Report
- Information Security Policy awareness http://www.informationshield.com

About the Author

Rajesh Ravichandran is Business Analyst at NIIT Technologies Ltd. He has over 5 years of experience in the airline industry. His expertise includes airline business know-how with specialization in passenger solutions and services; complex business analysis; requirement gathering and gap analysis.

About NIIT Technologies

NIIT Technologies is a leading IT solutions organization, servicing customers in North America, Europe, Asia and Australia. It offers services in Application Development and Maintenance, Enterprise Solutions including Managed Services and Business Process Outsourcing to organisations in the Financial Services, Travel & Transportation, Manufacturing/Distribution, and Government sectors. With employees over 7,000 professionals, NIIT Technologies follows global standards of software development processes.

Over the years the Company has forged extremely rewarding relationships with global majors, a testimony to mutual commitment and its ability to retain marquee clients, drawing repeat business from them. NIIT Technologies has been able to scale its interactions with marquee clients in the BFSI sector, the Travel Transport & Logistics and Manufacturing & Distribution, into extremely meaningful, multi-year "collaborations.

NIIT Technologies follows global standards of development, which include ISO 9001:2000 Certification, assessment at Level 5 for SEI-CMMi version 1.2 and ISO 27001 information security management certification. Its data center operations are assessed at the international ISO 20000 IT management standards.

NIIT technologies

India

NIIT Technologies Ltd.
Corporate Heights (Tapasya)
Plot No. 5, EFGH, Sector 126
Noida-Greater Noida Expressway
Noida – 201301, U.P., India
Ph: + 91 120 7119100
Fax: + 91 120 7119150

Americas

NIIT Technologies Inc.,
1050 Crown Pointe Parkway
5th Floor, Atlanta, GA 30338, USA
Ph: +1 770 551 9494
Toll Free: +1 888 454 NIIT
Fax: +1 770 551 9229

Europe

NIIT Technologies Limited
2nd Floor, 47 Mark Lane
London - EC3R 7QQ, U.K.
Ph: +44 20 70020700
Fax: +44 20 70020701

Singapore

NIIT Technologies Pte. Limited
31 Kaki Bukit Road 3
#05-13 Techlink
Singapore 417818
Ph: +65 68488300
Fax: +65 68488322