IBM InfoSphere Optim Data Masking solution

Highlights:

- Safeguard personally identifiable information, trade secrets and other sensitive data
- Easily mask confidential data using predefined transformations and site-specific routines
- Discover hidden instances of private data so that they can be fully protected
- Support compliance with privacy regulations and corporate governance standards

Today's organizations realize that data is a critical enterprise asset, so protecting that data—and the applications that hold that data—makes good business sense. However, different types of information have different protection and privacy requirements. Therefore, organizations must take a holistic approach to protecting and securing their business-critical information:

Discover where the data exists: You can't protect sensitive data unless you know where it resides across your enterprise and how it is related.

Safeguard sensitive data, both structured and unstructured: Structured data contained in databases must be protected from unauthorized access. File-level data encryption helps make this information unusable or unviewable except for those with specific rights. Unstructured data in documents and forms requires privacy policies to redact (remove) sensitive information while still allowing needed business data to be shared.

Secure and continuously monitor access to the data: Enterprise databases require real-time insight to ensure data access is protected and audited. Policy-based controls are required to rapidly detect unauthorized or suspicious activity and alert key personnel.

Protect non-production environments: Data in non-production, training and quality assurance environments needs to be protected against revealing sensitive information inadvertently yet must still be in usable form during the application development, testing and training processes.

By employing a data protection strategy across all areas and all types of data, organizations can ensure enterprise data is kept secure and protected.

Data privacy: The "untold story"

Data privacy protection is a tremendous focus for the IT community today. Organizations are making great strides to protect sensitive data in live application environments. But the "untold story" of implementing protection strategies in non-production (testing, development and training) environments remains a critical risk. As data breach headlines continue to mount, organizations must begin to address the most vulnerable areas of IT infrastructure—non-production environments.

So, what makes non-production environments so unique? The answer lies in the methods used to create non-production databases. Commonly, live production systems are cloned (copied) to a test environment—confidential data and all. Developers and QA testers find it easy to work with live data because it produces test results that everyone can understand. But do nonproduction environments actually require live data? The answer is no. Using realistic data is essential to testing, but live data values are not specifically necessary. Capabilities for "de-identifying" or masking production data offer a best practice approach for protecting sensitive data while supporting the testing process.

Data masking offers a best-practice approach

Data masking is the process of systematically transforming confidential data elements such as trade secrets and personally identifying information (PII) into realistic but fictionalized values. Data that has been scrubbed or cleansed in such a manner is considered acceptable to use in non-production environments. Masking enables developers and QA testers to use "production-like" data and produce valid test results, while still complying with privacy protection rules.

Data masking represents a simple concept, but it is technically challenging to execute. Most organizations operate within complex, heterogeneous IT environments consisting of multiple, interrelated applications, databases and platforms. IT managers do not always know where confidential data is stored or how it is related across disparate systems. The ideal solution must both discover sensitive data across related data stores and mask it effectively.

The IBM® InfoSphere™ Optim™ Data Masking solution provides comprehensive capabilities for masking sensitive data effectively across non-production environments, while still providing realistic data for use in development, testing or training. When you use InfoSphere Optim to mask confidential data, you protect privacy and safeguard shareholder value.

Implement proven data masking techniques

With InfoSphere Optim Data Masking solution, users can apply a variety of proven data transformation techniques to replace sensitive real data with contextually accurate and realistic fictitious data. Users can mask data in a single database or across multiple related systems. Simple examples of the masking techniques in InfoSphere Optim include substrings, arithmetic expressions, random or sequential number generation, date aging and concatenation. And the solution's context-aware masking capabilities help ensure that masked data retains the look and feel of the original information.

Those capabilities make it easy to de-identify many types of sensitive information, such as birth dates, bank account numbers, street address and postal code combinations and national identifiers (like Canada's Social Insurance numbers or Italy's Codice Fiscale).

The IBM InfoSphere Optim Transformation Library routines enable accurate masking of complex data elements, such as credit card numbers and email addresses. You can also incorporate site-specific data transformation routines that integrate processing logic from multiple related applications and databases. InfoSphere Optim offers the flexibility to support even the most complex data masking requirements.

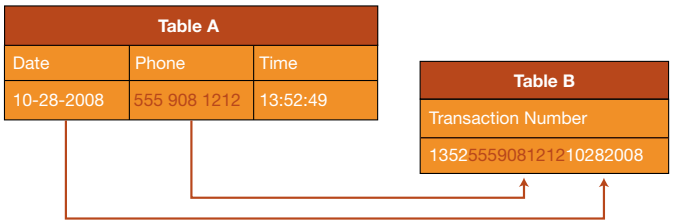

Figure 1: Confidential information hidden in compound fields poses a privacy risk to the organization.

Discover sensitive data

Some sensitive data is easy to find. For instance, credit card numbers in a column named "credit_card_num" will be simple to recognize. Most application databases, though, are more complex. Sensitive data is sometimes compounded with other data elements or buried in text or comment fields. Subject matter experts can sometimes offer insight, but only if they fully understand the system.

Figure 1 illustrates an example. Table A contains telephone numbers in the "Phone" column. In Table B however, the telephone number is obscured within a compound field in the "Transaction Number" column. Both instances represent confidential information that must be protected. But while data analysts can clearly recognize the telephone number in Table A, they may well overlook it in Table B. And every missed occurrence of private information represents a risk to the organization. What is the alternative?

Finding and masking data is part of the solution. There is an added complication. You need the capability to propagate masked data elements to all related tables in the database and across databases to maintain referential integrity. For example, if a masked data element, such as a telephone number, is a primary or foreign key in a database table relationship, then this newly masked data value must be propagated to all related tables in the database or across data sources. If the data is a portion of another row's data, it must be updated with the same data as well.

To minimize risk, data should be masked as close to its source system as possible. In some scenarios, data for tests is copied directly from a live system. In this case, data must be masked "in place" to ensure that the newly created test database is protected for use. In other scenarios, specific subsets of data are extracted using test data management products like the IBM InfoSphere Optim Test Data Management solution. In Figure 2, data is masked during the extract process to ensure that private information is never exposed.

Help ensure data integrity

IBM InfoSphere Discovery enables organizations to identify all instances of confidential data across the environment—whether clearly visible or obscured from view. InfoSphere Discovery works by examining data values across multiple sources to determine the complex rules and transformations that may hide sensitive content. It can locate confidential data items that are contained within larger fields, as described in the prior example, or that are separated across multiple columns. InfoSphere Discovery delivers automated capabilities that offer greater accuracy and reliability than manual analysis. When used together, the InfoSphere Optim Data Masking solution and InfoSphere Discovery provide the most effective, enterprise-scale solution for locating and masking sensitive data across complex, heterogeneous environments.

InfoSphere Discovery not only discovers hidden sensitive data, it also provides a full range of data analysis capabilities to discover hidden relationships and bring them clearly into view. By leveraging the combination of InfoSphere Discovery and the InfoSphere Optim Data Masking solution, all relationships will be uncovered and replacement values will be masked consistently and accurately across multiple data sources.

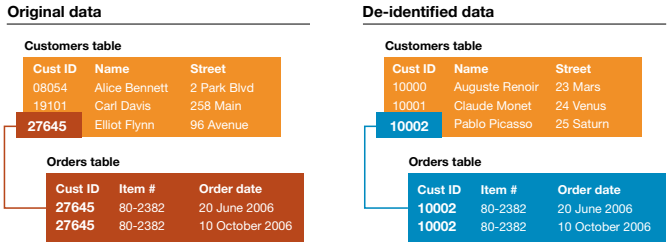

Figure 2: Data masking protects the confidentiality of private information and propagates it accurately throughout the system.

Support compliance initiatives

To support industry, government and internal compliance initiatives, data masking is a must. The European Union has established the Personal Data Protection Directive as the framework for privacy protection governing its member countries. And many other countries have similar regulations around the world. The U.S. Department of Health and Human Services has enacted the Health Insurance Portability and Accountability Act of 1996 (HIPAA), which addresses the privacy of individually identifiable health information. Additionally, industry coalitions are developing sector-specific governance standards, such as the Payment Card Industry Data Security Standard (PCI DSS), initiated by Visa and MasterCard. Implementing InfoSphere Optim helps you comply with these data privacy regulations by protecting the confidentiality of sensitive information across your enterprise.

InfoSphere Optim provides a scalable data masking solution with flexible capabilities that can be easily adapted to your current and future requirements. You also benefit from knowing that InfoSphere Optim supports all leading enterprise databases and operating systems, including IBM DB2®, Oracle, Sybase, Microsoft® SQL Server, IBM Informix®, IBM IMS™, IBM Virtual Storage Access Method (VSAM), Teradata, Adabas, Microsoft Windows®, UNIX®, Linux® and IBM z/OS®. In addition to providing data management support for all custom and packaged applications, InfoSphere Optim has the meta-model knowledge to support the key enterprise resource planning (ERP) and customer relationship management (CRM) applications in use today: SAP, Oracle E Business Suite, PeopleSoft Enterprise, JD Edwards EnterpriseOne, Siebel and Amdocs CRM.

About IBM InfoSphere

IBM InfoSphere Optim Data Masking solution is a key part of the InfoSphere portfolio. IBM InfoSphere software is an integrated platform for defining, integrating, protecting and managing trusted information across your systems. The InfoSphere platform provides the foundational building blocks of trusted information, including data integration, data warehousing, master data management and information governance, all integrated around a core of shared metadata and models. The portfolio is modular, allowing you to start anywhere, and mix and match InfoSphere software building blocks with components from other vendors, or choose to deploy multiple building blocks together for increased acceleration and value. The InfoSphere platform offers an enterprise-class foundation for information-intensive projects, providing the performance, scalability, reliability and acceleration needed to help simplify difficult challenges and deliver trusted information to your business faster.

For more information

To learn more about IBM InfoSphere, contact your IBM sales representative or visit: **ibm.com**/data/infosphere

To learn more about the IBM InfoSphere Optim Data Masking solution, please contact your IBM sales representative or visit: **ibm.com**/software/data/optim/protect-data-privacy

IBM Software Group
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
October 2010

Please Recycle

IMS14076-USEN-01