IBM InfoSphere Optim Data Masking solution

Mask data on demand to protect privacy across the enterprise

Highlights:

- Safeguard personally identifiable information, trade secrets, financials and other sensitive data
- Easily mask data on demand using predefined transformations and site-specific routines
- Respond in real time to suspicious requests for data
- Ensure a valid business need to know for sensitive data
- Discover hidden instances of private data so 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 it 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:

- **Understand where data exists:** Organizations can't protect sensitive data unless they know where it resides and how it's related across the enterprise.
- **Safeguard sensitive data, both structured and unstructured:** Structured data contained in databases must be protected from unauthorized access using data transformation techniques such as masking or encryption. Unstructured data in documents, forms, image files, GPS systems and more requires privacy policies to de-identify or mask sensitive data while still allowing needed business information to be shared.
- **Protect nonproduction environments:** Data in nonproduction, development, training and quality assurance environments needs to be de-identified or masked, yet still usable during the application development, testing and training processes.
- **Secure and continuously monitor access to the data:** Enterprise databases, data warehouses, file shares and Apache Hadoop-based systems require real-time monitoring and policies to ensure data access is protected and audited. Policy-based controls (like masking or connection termination) based on access patterns are required to rapidly detect unauthorized or suspicious activity and alert key personnel. In addition, data sources need to be protected against new threats or other malicious activity and continually monitored for weaknesses.
- **Demonstrate compliance to pass audits:** It's not enough to develop a holistic approach to data security and privacy. Organizations must also demonstrate and prove compliance to third-party auditors.

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 across the enterprise

News headlines about the increasing frequency of stolen information and identity theft have focused awareness on data privacy breaches and their consequences. Protecting data privacy is no longer optional—it's the law.

Organizations must have procedures in place to protect privacy in databases, applications and reports in both production and nonproduction systems to comply with data privacy regulations and avoid risk. As data-breach headlines continue to mount, it is clear that data is the most vulnerable enterprise asset.

Organizations need to adopt a policy-driven, on-demand masking approach to proactively protect data privacy and support compliance, especially in a computing era where data is everywhere and growing in volume, variety and velocity.

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. Masking enables receipts of the data to use "production-like" information while ensuring compliance 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. Organizations 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 sources and mask it effectively.

The IBM® InfoSphere® Optim™ Data Masking solution provides comprehensive capabilities to mask sensitive data effectively across applications, reports and databases in production and nonproduction environments. The InfoSphere Optim Data Masking solution de-identifies data anywhere a contextually accurate, yet fictionalized value is appropriate. For example, mask data in flight to fend off a hacker, mask data onscreen in a call center to ensure only those with a valid business need see sensitive client data, mask data in development, Q/A or testing environments, or mask data in extract, transform, load (ETL) or data movement solutions. When you use InfoSphere Optim to mask confidential data, you protect privacy and safeguard shareholder value.

The InfoSphere Optim Data Masking solution brings flexibility, scalability and adaptability to data masking by helping organizations:

- Understand where sensitive data exists
- Leverage masking services to mask data on demand, anywhere at any time
- Mask data in databases, warehouses and big data environments
- Mask data in both production and nonproduction environments
- Mask data on demand in applications or business reports to support real-time decision making
- Mask data on demand in the cloud
- Mask data in data movement tools such as ETL or data unload utilities

Proven data masking techniques

With the 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, across multiple related systems or in applications and reports. Simple examples of the masking techniques in InfoSphere Optim include substrings, arithmetic expressions, random or sequential number generation, date aging and concatenation. Plus, 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 (such as Canada's Social Insurance numbers or Italy's Codice Fiscale).

The IBM InfoSphere Optim Transformation Library routines are open and modular services enabling accurate masking of complex data elements, such as credit card numbers and email addresses on demand. 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.

The InfoSphere Optim Data Masking solution provides masking services to allow users to mask data on demand to meet business and compliance requirements (see Figure 1). Real-time capabilities to de-identify sensitive data across the enterprise will provide more flexible privacy protection in applications, databases, reports and more. The goal is to deliver integration and scalability as organizations embrace a new era of computing.

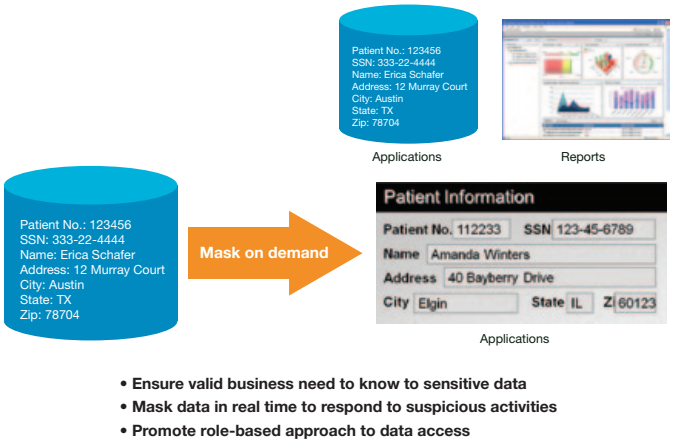

Figure 1: Mask data on demand.

Discovery of sensitive data

Some sensitive data is easy to find. For instance, credit card numbers in a column named "credit_card_num" are not difficult 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 2 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?

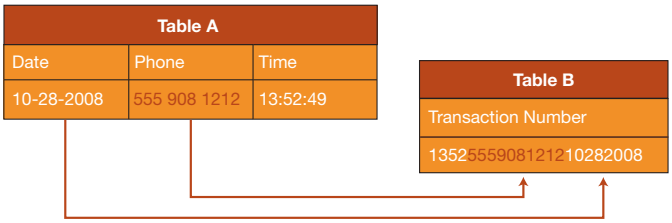

Figure 2: Confidential information hidden in compound fields poses a privacy risk to the organization.

Finding and masking data is part of the solution, but 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, that row 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 3, data is masked during the extract process to ensure that private information is never exposed.

Ensuring data integrity

IBM InfoSphere Discovery enables organizations to identify all instances of confidential data—whether clearly visible or obscured—throughout the environment. 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.

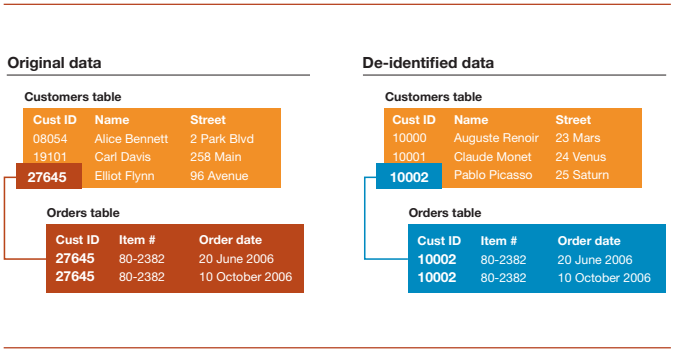

Figure 3: Data masking protects the confidentiality of private information and propagates it accurately throughout the system.

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.

Support for 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 US 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, IBM Netezza®, 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**/software/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 Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
October 2012

IMS14076-USEN-02