INFORMATICA

WHITE PAPER

The Informatica Solution for Data Privacy

Enforcing Data Security in the Era of Big Data

This document contains Confidential, Proprietary and Trade Secret Information ("Confidential Information") of Informatica Corporation and may not be copied, distributed, duplicated, or otherwise reproduced in any manner without the prior written consent of Informatica.

While every attempt has been made to ensure that the information in this document is accurate and complete, some typographical errors or technical inaccuracies may exist. Informatica does not accept responsibility for any kind of loss resulting from the use of information contained in this document. The information contained in this document is subject to change without notice.

The incorporation of the product attributes discussed in these materials into any release or upgrade of any Informatica software product—as well as the timing of any such release or upgrade—is at the sole discretion of Informatica.

Protected by one or more of the following U.S. Patents: 6,032,158; 5,794,246; 6,014,670; 6,339,775; 6,044,374; 6,208,990; 6,208,990; 6,850,947; 6,895,471; or by the following pending U.S. Patents: 09/644,280; 10/966,046; 10/727,700.

This edition published June 2012

Table of Contents

Executive Summary

Data breaches in the enterprise are a disturbingly common occurrence, with companies and government agencies suffering from them almost daily. In addition to negative publicity, these breaches often have far-reaching effects, including such costs as regulatory fines, litigation fees, consulting fees, and the loss of customers. As the Ponemon Institute has reported, the average cost of a data breach in the United States reached $5.5 million dollars in 2011.[1]

Big data is inherently vulnerable to these breaches, making the matter of preventing them all the more urgent. But doing so has become particularly challenging due to the increasingly complex IT environment. As a result, organizations are in greater need of a robust data privacy solution to prevent breaches and enforce data security. Such a solution should empower IT organizations to:

- Protect sensitive data exposed in production and nonproduction environments
- Minimize the performance impact of application source code changes
- Respond quickly to reduce costs and risks
- Preempt and prevent data breaches

The Informatica® solution for data privacy is based on proven technology that substantially reduces the risk of a data breach and helps organizations comply with data privacy policies, regulations, and mandates at lower costs. The solution consists of three products:

1. Informatica Data Subset
2. Informatica Persistent Data Masking
3. Informatica Dynamic Data Masking

Together these products mask or block sensitive and confidential information from unauthorized access in production and nonproduction systems. They enable application end users, database administrators, developers, testers, trainers, production support, and business analysts to perform their functions without sacrificing data security.

What makes the Informatica solution for data privacy unique is that it is based on the industry-leading Informatica Platform. This comprehensive, open, unified, and economical data integration platform supports a centralized data management approach, so your IT organization can leverage the solution across multiple business lines to conduct audits and comply with data privacy mandates enterprise-wide.

This white paper explains how the Informatica solution for data privacy helps organizations to meet the challenges of enforcing data security in the era of big data. It also discusses how the solution supports data governance programs, empowering IT organizations to ensure that data privacy is not just a one-time initiative, but part of an overall, ongoing data governance program.

[1] Ponemon Institute, 2012 Annual Study: Global Cost of a Data Breach.

Data Breaches in the Era of Big Data

By virtue of its volume, variety, and velocity, big data is particularly prone to data breaches. Big data is the confluence of three technology trends:

- The massive growth of transaction data volumes
- The explosion of new types of interaction data, such as social media and device data
- Highly scalable data processing with Hadoop

In a study by the research firm IDC, it was estimated that the amount of information managed by enterprise data centers will grow by a factor of 50 over the next decade.[2] In addition, social media, mobile computing, and device-to-device interactions have increased the variety of data types available to the enterprise for data analytics. Finally, the sheer speed at which data moves from point to point is forcing organizations to take a new look at how they approach data privacy across the enterprise and decide whether sensitive data should reside at the front, middle, or back of office applications. As big data volumes grow, these new strategies need to be able to scale along with them.

The Ponemon Institute study found that 42 percent of data breaches are caused by insider negligence. Even as organizations set up sophisticated barriers to protect themselves from external threats, there often remains the risk of internal threats that are just as dangerous. A similar study conducted by Verizon Communications found that 50 percent of data breaches involved database or application servers and if stolen, 98 percent of the records contained therein were breached.[3] Unlike with an end-user device such as a USB thumb-drive, nearly all the records of a database are vulnerable to theft if breached.

Preventing data breaches becomes even more challenging due to the increasingly complex IT environment. Most IT organizations need to develop and maintain multiple applications to support individual business units. For each production application, there may be multiple copies for patch, test, development, and training purposes—not to mention backup or remote copies to support data protection and disaster recovery strategies. Each of those copies, in turn, may have a number of resources with direct access to systems containing data that's potentially sensitive or subject to privacy regulations. To mitigate the escalating costs that often result from this complexity, IT organizations in many cases either employ offshore resource models or deploy software as a service (SaaS) or cloud-based offerings.

[2] IDC, "The 2011 Digital Universe Study: Extracting Value from the Chaos," June 2011.

[3] 2010 Data Breach Investigation Report, Verizon Risk Team in conjunction with United States Secret Service.

Organizations need a robust data privacy solution to prevent data breaches and enforce data security in today's complex IT environment. The solution should empower IT organizations to:

1. **Protect sensitive data exposed in less secure, nonproduction environments.** It is common practice to utilize real production data in development, testing, and training activities by making full-size copies of that data. Typically, IT organizations will create anywhere from 6 to 12 copies of production data sets. This practice, however, exposes sensitive data to privileged, knowledgeable, sophisticated, and IT-savvy individuals who may not have the best intentions in mind. A solution that instead provides testers with a sanitized version of production data gives them the confidence of having quality data sets while alleviating concerns of potential data theft from within the less-secure, nonproduction environments.

2. **Secure production data.** Increasingly, companies are outsourcing application development and support. IT organizations need to ensure that sensitive data is secure in the application screens as well as in the database for the privileged users or database administrators. Production data must be protected from unauthorized access by production support teams, outsource personnel, developers, and database administrators. The challenge is to protect data while not impacting application functionality.

3. **Minimize the performance impact of application source code changes.** One way to address data privacy issues is to implement whole-database, table-level, or column-level encryption or tokenization. This approach utilizes role-based access to enforce data privacy at the application tier where personal information could otherwise be exposed. The challenge with encryption and tokenization is that both typically require source code changes, which could potentially impose a performance overhead. Additionally, because developers would be required to implement the code changes, it makes enforcing segregation of duties much more difficult. An ideal solution protects data with minimal or no performance impact and minimal or no application code changes.

4. **Respond quickly to reduce costs and risks.** As data security regulations increase and change, IT organizations need to respond quickly. Having a solution that can be rapidly implemented and flexibly scaled to support new and evolving regulations greatly reduces costs and risks. It also cuts down on the time required to respond to a new audit request. Any time a new regulation is mandated, IT organizations should be able to deploy the new security policy quickly and easily.

5. **Preempt and prevent data breaches.** In the past, detailed audits were conducted only when personal information was breached. Today's audits are more thorough. To mitigate the risk of data breaches while complying with increasingly strict data security regulations, IT organizations need to implement a solution for preventing data breaches as well as for proactively identifying where breaches may occur in the future.

The Informatica Solution for Data Privacy

The Informatica solution for data privacy is based on proven technology that substantially reduces the risk of a data breach and helps organizations comply with data privacy regulations. The solution consists of three products:

1. Informatica Data Subset
2. Informatica Persistent Data Masking
3. Informatica Dynamic Data Masking

Together these products mask or block sensitive and confidential information from unauthorized access in production and nonproduction systems, reducing the risk of data breaches. They empower your IT organization to comply with data privacy policies, regulations, and mandates at lower costs. They enable your application end users, database administrators, developers, testers, trainers, production support, and business analysts to perform their functions without sacrificing data security.

Let's dive deeper into each product in the solution.

Informatica Data Subset

Informatica Data Subset is flexible, scalable software for creating data subsets. It enables your IT team to create, update, and secure data subsets—smaller, targeted databases—from large, complex databases. With referentially intact subsets of production data, your IT organization dramatically reduces the amount of time, effort, and disk space needed to support nonproduction systems.

Informatica Data Subset quickly replicates and refreshes production data with only the most relevant, high-quality application data. This means your IT team doesn't need to create a full database copy—you separate out only functionally related data from interconnected systems.

As you create subsets, you can apply data masking policies so that the data in your smaller production copy for testing purposes is secure. Informatica Persistent Data Masking enables to you to create and maintain these global masking rules.

Informatica Persistent Data Masking

Informatica Persistent Data Masking is scalable data masking software that helps your entire IT organization manage access to your most sensitive data. The software shields confidential data, such as credit card information, Social Security numbers, names, addresses, and phone numbers, from unintended exposure to reduce the risk of data breaches. It provides unparalleled enterprise-wide scalability, robustness, and connectivity to a vast array of databases.

Informatica Persistent Data Masking minimizes the risk of data breaches by masking test and development environments created from production data regardless of database, platform, or location. The software provides sophisticated, but flexible, masking rules that allow your IT team to apply different types of masking techniques to various data used in testing, training, and other nonproduction environments.

With Informatica Persistent Data Masking, IT organizations can create enterprise-wide data privacy polices while maintaining a segregation of duties. Auditors and security officers can define policies while developers, testers, and trainers retain access to contextually rich, functionally intact, and realistic looking data without impacting application functionality.

Informatica Data Subset and Informatica Persistent Data Masking together form Informatica Test Data Management (see Figure 1). This solution enables security administrators, QAs, and other users to discover where sensitive data lives across the organization, make subsets of production data to create smaller copies for testing or training purposes, mask the data, and then substantiate that the data was masked according to specified privacy policies.

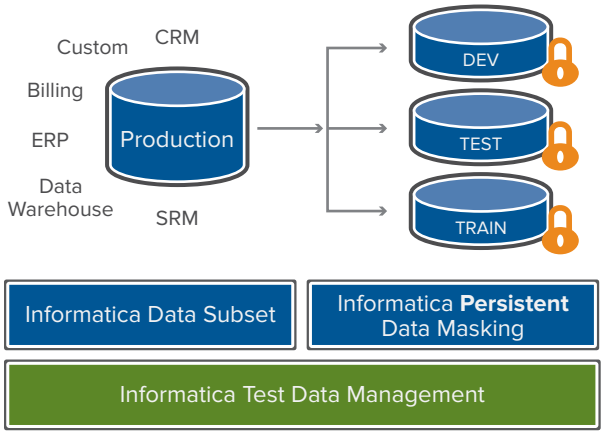


Figure 1: With the Informatica Test Data Management solution, your IT team can apply different subset and masking techniques to production data. The newly secure data can be used in testing, training, and other nonproduction environments.

Informatica Dynamic Data Masking

In production environments, privileged users such as DBAs or functional business users often have inadvertent access to sensitive data that they don't actually need to perform their jobs. As an example, a DBA might require the use of a production billing system to examine performance issues. In that scenario, there would be no need for the DBA to see sensitive data such as customer credit information.

To help organizations proactively address this data privacy challenge in production, Informatica offers Informatica Dynamic Data Masking—the only true dynamic data masking product on the market. Informatica Dynamic Data Masking de-identifies data and controls unauthorized access to production environments. The software dynamically masks sensitive information in production data and blocks, audits, and alerts end users, IT personnel, and outsourced teams who access sensitive information (see Figure 2).

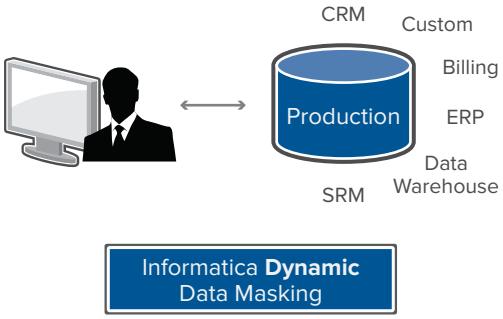


Figure 2: Informatica Dynamic Data Masking dynamically masks sensitive information in production data, and blocks, audits, and alerts end users to its unauthorized access.

With Informatica Dynamic Data Masking, your IT organization can apply sophisticated, flexible data masking rules based on a user's authentication level. Through a simple yet elegant rules engine, criteria can be specified to identify which SQL statements are to be acted upon (rewritten). When there is a match, Informatica Dynamic Data Masking applies one or more actions—including mask, scramble, hide, rewrite, block, or redirect—to prevent unauthorized users from accessing sensitive information in real time.

The Informatica Advantage

What makes the Informatica solution for data privacy unique? It is based on the Informatica Platform, the industry-leading data integration platform. This comprehensive, open, unified, and economical data integration platform supports a centralized data management approach, so your IT organization can leverage the solution across multiple business lines to conduct audits and comply with data privacy policies and regulations enterprise-wide.

The Informatica solution for data privacy supports your organization's data governance program (see Figure 3). This comprehensive solution empowers your IT organization to ensure that data privacy is not just a one-time initiative, but part of an overall, ongoing data governance program by:

- Addressing all database applications, on or off premises, including both production and nonproduction databases
- Providing a centralized management and control center for consistent enterprise-wide data privacy protection
- Featuring coarse and fine-grained encryption to support a variety of custom and packaged applications, databases, and data center policies
- Handling data volume growth—either organic growth or as new applications are deployed in the data center

Leveraging the Informatica Platform, the Informatica solution for data privacy enables you to address each part of the lifecycle:

1. **Define and classify** the sensitive data of the organization.
2. **Discover** where that sensitive data lives across databases and applications.
3. **Apply** policies of creating subsets and masking consistently across the systems of an organization to meet various compliance standards.
4. **Measure and monitor** and then prove that the data has been protected.

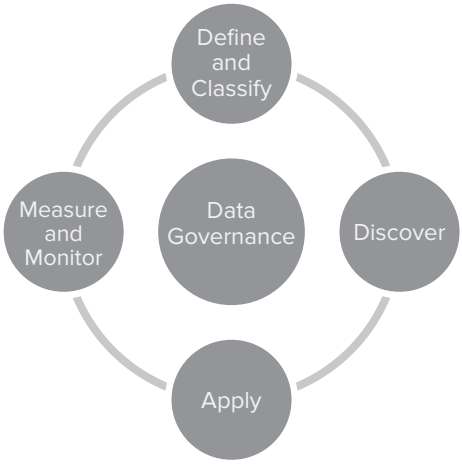


Figure 3: The Informatica solution for data privacy supports data governance best practices.

Let's examine how the Informatica solution for data privacy handles five key tasks:

- Protects sensitive data exposed in less secure nonproduction environments
- Secures production data
- Minimizes the performance impact of application source code changes
- Accelerates deployment to reduce costs and risks
- Preempts and prevents data breaches

Protects Sensitive Data in Less Secure Nonproduction Environments

Data encryption solutions perform a valid function, but they do not prevent access by standard IT users or when authenticated applications and tools access the databases. In these cases, all values are returned in the clear.

Format preserving encryption is well suited for encrypting formatted data items, such as account numbers, Social Security numbers, and credit card numbers, and doing so in a reversible manner. It does not, however, protect data of variable length or unstructured data such as documents, spreadsheets, and transactions in SWIFT, NACHA, or other semistructured formats.

The Informatica solution for data privacy ensures that access is restricted to only DBAs, system administrators, and other privileged users as needed. The solution protects unstructured and semistructured data of variable length. It ensures that data is protected without source code or database changes. It is also ideal for testing, training, and quality assurance environments in which data is protected as the relevant environment is created.

Secures Production Data

Database activity monitoring enables organizations to monitor the who, what, when, and where of general database activity and access. Database activity monitoring sits between a database management system and the target database to collect information about any connection between the two and then highlights where breaches could occur. Yet normally during this process, private information is neither hidden nor secured.

By augmenting database activity monitoring with data masking, the Informatica solution for data privacy ensures that sensitive data is protected from end users, IT, and outsourced staff. The solution can mask, scramble, hide, or apply row/column level security measures to personal information. In addition, it provides full integration to ActiveDirectory/LDAP. In this way, all data is protected according to the rights and privileges of the user.

Minimizes the Performance Impact of Application Source Code Changes

Tokenization is well known for supporting PCI data and preserving the format and width of a table column. But because many tokenization solutions today require creating database views or changing application source code, they are not compliant with packaged applications that don't allow these changes. In addition, databases and applications take a measurable performance hit to process tokens.

The Informatica solution for data privacy complements tokenization and encryption solutions because it doesn't require changes to the database or the application. For nonproduction environments, it provides data protection without the performance hit to process the tokens. And for production environments, it dynamically restricts access to privileged users such as DBAs and system administrators.

Accelerates Deployment to Reduce Costs and Risks

The Informatica solution for data privacy accelerates deployments through a purpose-built interface that makes it easy to create and manage global rules. In this central environment, administrators can:

- Choose from existing masking algorithms (e.g., blurring, substitution, shuffling) to use in their global privacy rules
- Use prebuilt accelerators for PeopleSoft, Oracle, and SAP to deploy data subset strategies quickly and easily without the fear that subset policies might cease to function when the underlying application changes
- Use predefined rules to quickly and dynamically protect packaged applications, such as SAP, PeopleSoft, Oracle E-Business, and Siebel
- Accelerate deployments and quickly secure complex applications with data packs for Personally Identifiable Information (PII), Protected Health Information (PHI), and Payment Card Industry Data Security Standard (PCI-DSS).

Preempts and Prevents Data Breaches

The Informatica solution for data privacy enables organizations to implement and easily maintain data privacy policies (see Figure 4) in support of data governance programs. It empowers organizations to both prevent and preempt data breaches by:

- Discovering and identifying where sensitive data lives in the organization
- Defining and classifying sensitive data, and setting data privacy policies appropriately and consistently across multiple applications and database instances
- Masking data according to all required rules (e.g., PHI, PII, and PCI-DSS)
- Applying data privacy policies to nonproduction environments by creating test data subsets of production data for nonproduction use cases (e.g., when test data sets are created from production data, masking sensitive data greatly reduces risks of exposure)
- Creating subsets of test data based on a single source of masked test data consistently, decreasing the risk of a data breach and complying with regulations
- Creating an audit trail to validate that data has been protected

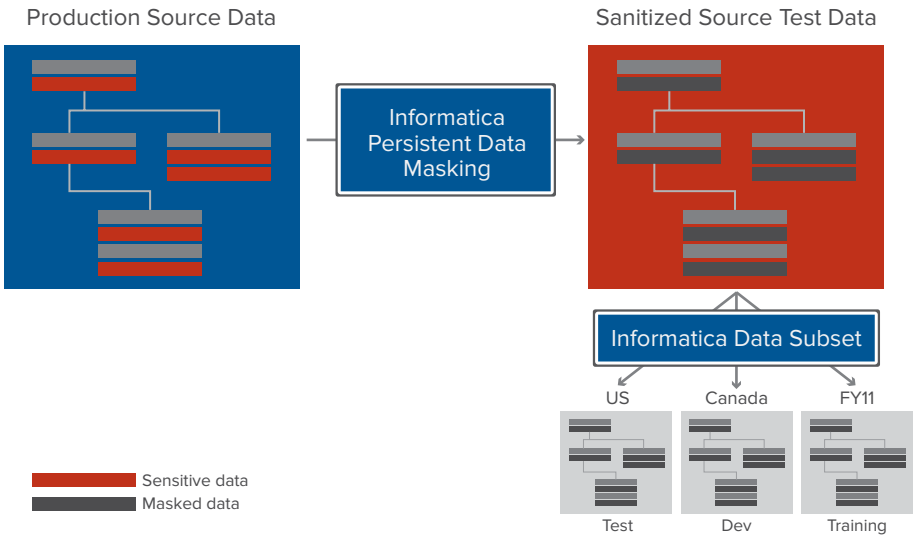


Figure 4: Informatica provides an end-to-end solution set for maintaining data privacy policies.

In contrast to application-specific industry solutions that are costly to maintain, the Informatica solution for data privacy supplies economies of scale by centralizing data privacy. The solution scales to support increasing volumes of data and connect to all types of systems across the many business units in need of support.

Conclusion

The Informatica solution for data privacy is based on proven, next-generation technology that empowers companies and government agencies to quickly, easily, and cost-effectively manage and ensure data privacy throughout their organizations. Utilizing an open architecture, the solution is highly scalable and flexible, enabling users to protect private and sensitive data, decrease the risk of data breaches, effectively meet compliance requirements for production and nonproduction environments on a timely basis, and lower the cost of data while increasing its value and maximizing its return.

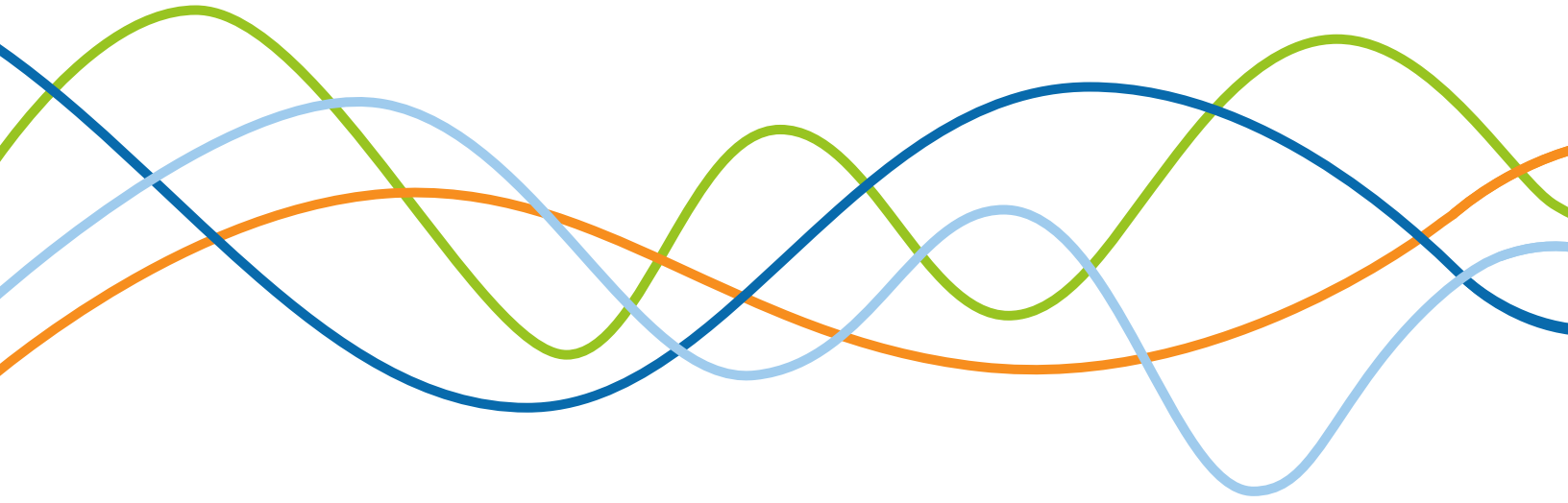
This data privacy solution offers the following benefits:

- Ability to define sensitive data, identify where it resides, dynamically or persistently mask it, and create functional and secure data subsets
- Complete transparency in data masking, whether by avoiding changes to production data or providing an audit trail of masked data
- Quick security to complex applications through predefined data privacy rules and an intuitive, powerful user interface for creating and maintaining new rules
- High-performance scalability to support large volumes of data, dynamically or in batch
- Comprehensive connectivity to databases, mainframes, and applications

With its unique advantages, the Informatica solution for data privacy helps organizations to meet the challenges of enforcing data security in the era of big data. Built on an industry-leading data integration platform, this solution supports a centralized data management approach, empowering IT organizations to ensure that data privacy is not just a one-time initiative but part of an overall, ongoing data governance program.

ABOUT INFORMATICA

Informatica Corporation (NASDAQ: INFA) is the world's number one independent provider of data integration software. Organizations around the world rely on Informatica for maximizing return on data to drive their top business imperatives. Worldwide, over 4,630 enterprises depend on Informatica to fully leverage their information assets residing on-premise, in the Cloud and across social networks.



INFORMATICA®

Worldwide Headquarters, 100 Cardinal Way, Redwood City, CA 94063, USA
phone: 650.385.5000 fax: 650.385.5500 toll-free in the US: 1.800.653.3871
informatica.com linkedin.com/company/informatica twitter.com/InformaticaCorp

© 2012 Informatica Corporation. All rights reserved. Printed in the U.S.A. Informatica, the Informatica logo, and The Data Integration Company are trademarks or registered trademarks of Informatica Corporation in the United States and in jurisdictions throughout the world. All other company and product names may be trade names or trademarks of their respective owners.

IN09_0612_02033