

Informatica PowerCenter Data Masking Option

Protecting Sensitive, Private Information

Christophe Pichot-Meuger

Architect

Cegedim

“As a company dedicated to giving organizations enterprise-wide access to data, Informatica has been mindful of the mandate to keep that data secure. With PowerCenter 8.5, they’ve given us additional ICC-grade security features.”

Organizations take extreme measures to secure private data in production environments. But often nonproduction environments, such as development, test, or training, are overlooked. These environments require realistic data and are often provided with a copy of the production data. As a result, nonproduction environments are an attractive target to malicious users. Forrester Research recommends that “. . .all enterprises should look at data masking technology that can help protect private data in test environments or when the data is sent to an outsourcing or offshore vendor.”¹

The **PowerCenter® Data Masking Option** is used to protect sensitive, private information by masking it in-flight to produce deidentified but realistic-looking data. This option features multiple data masking techniques and algorithms to ensure randomization while maintaining the original nature of the data and preserving referential integrity. This option also includes specialized, built-in content and rules for common sensitive fields, such as name, address, Social Security number, credit card number, and telephone number.

Fully integrated within PowerCenter’s architecture, these data masking capabilities automatically leverage the core strengths of PowerCenter, including universal data access, rich transformations, high performance, and a robust security framework.

BENEFITS

- Reduce the risk of data and compliance breaches by protecting sensitive private data
- Decrease development costs by reducing the effort needed to mask data
- Accelerate time to market by increasing the quality of development and testing

Data Masking Option Key Features

Mask Sensitive, Private Information

Multiple techniques and algorithms to mask sensitive fields and produce realistic, fully-functional data, including:

- **Non-Deterministic Randomization.** Replace a sensitive field with a randomly generated value subject to various constraints

Example: Generate a random date between 01/01/1910 and 12/31/2010

- **Blurring.** Add a random variance to the original value

Example: Replace a sales amount with a random value but within a 5% range of the original value

¹ Source: Forrester Research, “Protecting Private Data With Data Masking”, March 2006.

- **Repeatable Masking.** Maintain referential integrity by generating values that are both repeatable and unique

Example: Replace the tax ID number 12-3454165 with 32-9843454 consistently

- **Substitution.** Randomly substitute original values with false but realistic-looking values

Example: Substitute “John Smith” with “Glen Carter”

Built-in methods maintain repeatability and preserve referential integrity across multiple tables and fine-grained controls produce randomized output while also preserving the original data properties.

Specialized, Built-In Rules and Content for Common Data Masking Scenarios

- Content for substituting name and address data
- Prepackaged rules for special fields such as Social Security number, credit card number, and telephone number

Example: Generate a random Social Security number that is structurally correct but not actually a valid number

Example: Generate a random credit card number that maintains the issuer ID and checksum validation rule

- Sample mappings that demonstrate common data masking scenarios

Full Integration Across Entire Data Integration Platform

- Leverages PowerCenter connectivity to mask data from virtually all enterprise data sources, such as relational, mainframe and unstructured
- Automatically harnesses all of PowerCenter’s rich transformation capabilities to meet any data masking need

- Leverages the high security standards of PowerCenter to provide robust security for the data masking solution

- Seamlessly integrates into existing PowerCenter environments through a simple plug-in

Data Masking Option Benefits

Decrease Risk of Data and Compliance Breaches by Protecting Sensitive Private Data

One of the common reasons for data and compliance breaches is the exposure of sensitive information in nonproduction environments such as development and test, which may be offshore or outsourced. The Data Masking Option reduces this risk by deidentifying sensitive, private information, and providing fully functional data for such environments. This option includes sophisticated techniques and algorithms that completely obfuscate the original data. Fully integrated within PowerCenter, this option leverages PowerCenter’s high security standards to provide enterprise-grade security for the data masking solution.

Reduce Development Costs by Reducing the Effort Needed to Mask Data

Creating deidentified data in nonproduction environments isn’t an easy task—especially when there are large databases with hundreds of interdependent tables. Forrester Research estimates that it takes anywhere from four to six months to create a complete data masking solution by hand. With this prepackaged option, organizations can reduce development efforts by utilizing the specialized, built-in rules and content for common data masking scenarios. Because this option is seamlessly

integrated within PowerCenter, developers can quickly come up to speed and also leverage built-in connectivity, reusability, and transformation capabilities for building the data masking solution.

Accelerate Time to Market by Increasing the Quality of Development and Testing

Project slippage often occurs when “dummy” data is used in the development and test environments. While usage of dummy data protects privacy, it doesn’t accurately reflect the real production data. The Data Masking Option features robust capabilities for creating fully-functional, production-like data that retains the original data’s inherent properties, such as format, width, and range. This helps IT teams develop and deliver successful applications that conform to specifications with high fidelity and accuracy, reducing project risk and accelerating time to market.

Learn More

Let Informatica help you determine how the PowerCenter Data Masking Option can maximize the business value of your data.

Visit us at www.informatica.com/powercenter or call 1.800.653.3871.

About Informatica

Informatica is a leading provider of enterprise data integration software and services. With Informatica, organizations can gain greater business value by integrating all their information assets from across the enterprise. Thousands of companies worldwide rely on Informatica to reduce the cost and expedite the time to address data integration needs of any complexity and scale.

INFORMATICA[®]
The Data Integration Company™

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 www.informatica.com

Informatica Offices Around The Globe: Australia · Belgium · Canada · China · France · Germany · Ireland · Japan · Korea · the Netherlands · Singapore · Switzerland · United Kingdom · USA

© 2007 Informatica Corporation. All rights reserved. Printed in the U.S.A. Informatica, the Informatica logo, and PowerCenter 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.