INFORMATICA

WHITE PAPER

Informatica Dynamic Data Masking

Preventing Data Breaches with Benchmark-Proven Performance

This edition published November 2012

Table of Contents

Executive Summary 2

Challenges to Securing Large Data Volumes 3

Securing Access to Sensitive Information by End Users 3

Securing Data for IT Staff, Production Support Teams, Third-Parties, and Outsourced Personnel 3

Informatica Meets the Challenges 4

Informatica Dynamic Data Masking 4

High Performance Through Design and Architecture 6

Informatica Dynamic Data Masking Performance Benchmark 6

Test Conditions 6

Test Methodology 7

Benchmark Results 7

Informatica Dynamic Data Masking in Action 8

Conclusion 9

Executive Summary

Public companies, private institutes, and government agencies often inadvertently expose company-confidential and private information to end users, IT staff, business support teams, and outsourced personnel. This introduces the risk of unauthorized access to that data. To prevent damage to their reputation, minimize unforeseen costs, and limit regulatory penalties, these organizations must aggressively invest in data security technologies. Their goal is to minimize the risk of data breaches and ensure compliance with privacy regulations by de-identifying data and controlling unauthorized access to data in both production and nonproduction environments.

Due to dramatic increases in the complexity of IT environments, the introduction of on-premise and public cloud architectures, and the resource demands of growing data volumes, big data has become especially vulnerable to these data breaches. This vulnerability has made the matter of preventing breaches all the more urgent. As a result, organizations are in greater need of high-performance, robust, and scalable data security software to prevent breaches and enforce compliance.

Informatica® Dynamic Data Masking 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. This white paper explains how the software's architecture is well suited to the unique challenge of protecting increasingly large volumes of data in a scalable manner. The paper then presents a benchmark measuring the software's highly efficient throughput and resource consumption, substantiating its top performance through design and architecture.

These benchmark results, along with an example of the software in action, illustrate the ability of Informatica Data Dynamic Masking to effectively secure sensitive information in real time—while enabling users to perform their jobs without disruption.

Challenges to Securing Large Data Volumes

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. Forrester has reported that 70 percent of data breaches are caused by insiders.[1] Similarly, in a Ponemon Institute report from May 2012, organizations surveyed said 50 percent of these cases involve a malicious insider such as a privileged user.[2]

Protecting large volumes of data is difficult due to the management complexities it introduces in IT environments, including numerous end points to which data gets distributed (e.g., PCs, mobile phones, tablets). To mitigate the escalating costs that often result from developing applications in this landscape, IT organizations in many cases employ teams of offshore resources, deploy software as a service (SaaS), or leverage cloud-based offerings. But these approaches introduce new risks of exposing sensitive information to both insiders and outsiders—especially when using existing data security approaches.

Securing Access to Sensitive Information by End Users

As business users analyze increasing volumes of data, the need to protect the sensitive information in that data only grows. End users require ever-increasing access controls and face restrictions imposed by privacy regulations on a global scale. Here are two examples:

- The human resources employee of a U.S.-based company is prohibited from viewing certain personal information details about his European peers
- A banker accessing her client's information from the Swiss office of her company must include additional restrictions when accessing the same application from the Paris office (this is known as location-based access control)

Conventional encryption solutions require transactions to be encrypted on every read and decrypted on every write. But this approach is problematic. An analyst in a U.S. bank trying to identify who might default on their loans has to query millions of rows of data, searching on numerous attributes such as loan date, loan amount, zip code and state where that loan was made, and the data of the last payment. If these sensitive details are encrypted and then decrypted on every query, the performance impact may bring down the entire database.

Securing Data for IT Staff, Production Support Teams, Third-Parties, and Outsourced Personnel

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. In fact, all regulations today mandate that access to sensitive and personal information should be on a "need to know" basis. But meeting this requirement becomes a challenge when internal teams performing their jobs need to access production environments but must not see the sensitive information within them. The difficulty lies in protecting data without impacting the application that houses it. The problem is exacerbated by the large volumes of data that need to be accessed on this "need to know" basis.

[1]Forrester, "Test Data Privacy Is Critical to Meet Compliances," October 2009.
[2]Ponemon, "Safeguarding Data in Production & Development: A Survey of IT Practitioners," May 2012.

Informatica Meets the Challenges

Informatica Dynamic Data Masking

As a result of these challenges, organizations are in greater need of robust data masking software to prevent breaches and enforce data security. Such a solution should empower IT organizations to:

- Mask the sensitive data exposed in production environments
- Shield production applications and databases without changes to source code
- Respond quickly to reduce the risks of data breaches and the resulting costs
- Customize database security for different regulatory or business requirements

Informatica Dynamic Data Masking helps organizations to accomplish these daunting tasks, proactively addressing data privacy challenges *in real time*. As the only true dynamic data masking product on the market, Informatica Dynamic Data Masking de-identifies data and controls unauthorized access to production environments (see Figure 1).

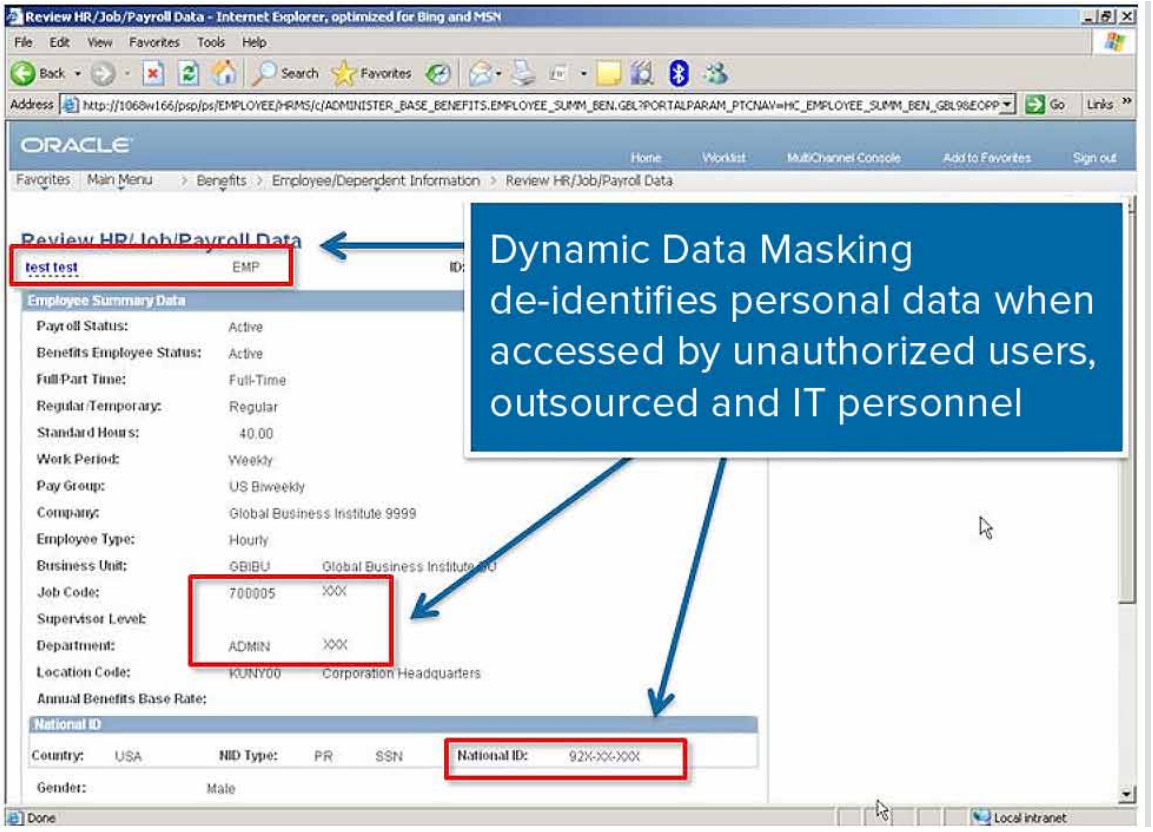


Figure 1: Informatica Dynamic Data Masking dynamically masks sensitive information in production data, blocks and audits unauthorized access to it, and alerts end users to that access.

Informatica Dynamic Data Masking is built on a patented database network in-line proxy, transparently installed between applications and databases. Acting as a database listener, the proxy processes all inbound application requests coming from application screens, canned reports, and DBA/development tools. Once they are analyzed or acted on, they are sent to the database for prompt execution. 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, the software applies one or more actions—including mask, scramble, hide, rewrite, block, or redirect—to prevent unauthorized users from accessing sensitive information in real time.

Informatica Dynamic Data Masking has no throughput effect on a production environment, supporting thousands of SQL requests per second. This is because it was designed for performance. With its multithreading capability, the software allows linear scalability to ensure support of the most demanding throughputs with the smallest footprint, averaging about 2 percent of the overall database server CPU resources (when installed on the database). Alternatively, Informatica Dynamic Data Masking can be installed on a dedicated server, which results in 0 percent database server resource overhead.

The high performance of Informatica Dynamic Data Masking has been proven, both in our high-load benchmarks—where it secures business applications with throughputs of more than 5,000 SQL requests per second while consuming less than 2 GigaHertz CPU processing—and in our quickly increasing install base, deploying real-time data masking in the most demanding business applications.

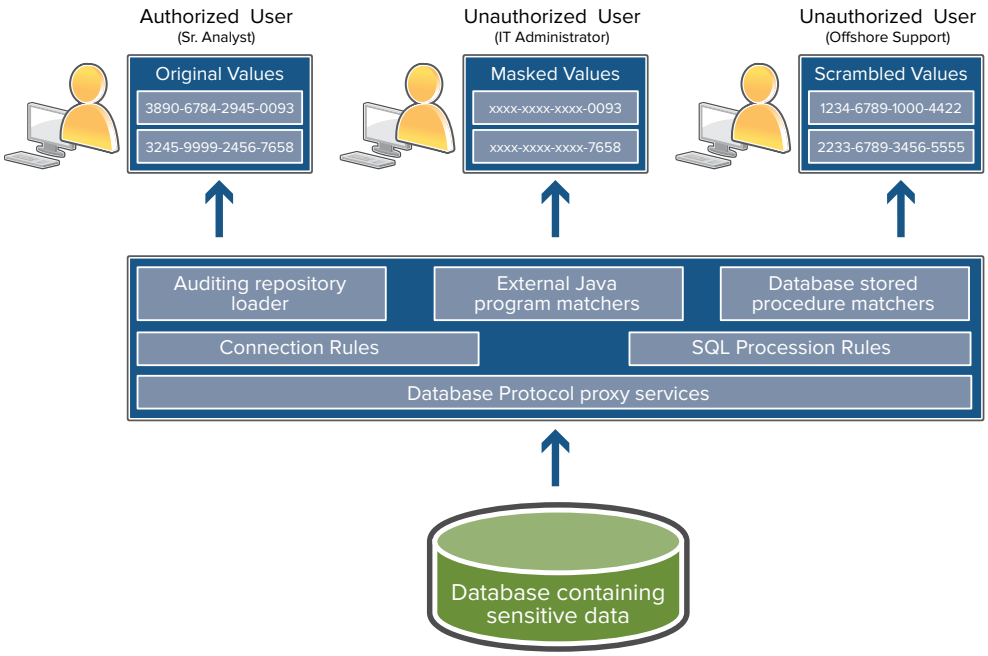


Figure 2: Informatica Dynamic Data Masking is built on a patented database network in-line proxy installed between applications and databases.

High Performance Through Design and Architecture

Informatica Dynamic Data Masking was designed from the start to optimize performance for databases in the following ways:

1. **Multithreading.** Informatica Dynamic Data Masking was designed and built using a multithreaded model. Using an adaptive model, DDM opens processing threads that handle incoming traffic according to the number of connections and the volume of SQL requests per second, thus ensuring linear scalability in its throughput.

2. **Selective application.** Conventional encryption solutions require transactions to be encrypted on every read and decrypted on every write. But Informatica Dynamic Data Masking secures access to information by rewriting only the very few "SELECT" requests that are submitted by unauthorized users (such as production support team, IT staff, and outsource personnel). This means the software masks only a fraction (<0.1 percent) of all application SQL request traffic, providing a pass-through for the majority of remaining SQL traffic.

3. **Batch bypassing.** Informatica Dynamic Data Masking has the internal switching capability to bypass internal batch jobs. Internal programs and ETL directly access the database, further reducing the amount of SQL statement traffic flowing through the software. With a large telecom customer, for example, Informatica Dynamic Data Masking bypasses billing batches, ETL extracts, and off-line billing processing to directly connect to the database. This reduces traffic evaluated by the software by more than 60 percent.

4. **SQL-based anonymity.** Informatica Dynamic Data Masking does not mask the data returning to the user. It makes the incoming SQL "SELECT'" list (SQL rewrite) anonymous, so no propagation is seen even when extracting millions of masked records. For example, when retrieving one million customer names, the database simply returns a substring of the names + "xxx" (for example, changing "Scott" to "Scxxx"). This results in zero impact to system performance. Because the software changes only the "SELECT" list (to a substring in this example), it skips over the "WHERE" clause or "GROUP BY/ORDER BY" clauses. Thus, existing database execution plans and response times remain unchanged.

5. **Seamless failover.** Informatica Dynamic Data Masking can be installed on a high-availability cluster, with automatic failover between the software's nodes and application failover onto the database. This means that there is an automatic and seamless failover because Informatica Dynamic Data Masking is treated like another application or server.

Informatica Dynamic Data Masking Performance Benchmark

Informatica has performed a benchmark that measured the throughput impact when installing Informatica Dynamic Data Masking to secure large high-transactional business applications and the overhead resource consumption on an existing database server. The software provides a highly scalable multithreaded SQL processing server that easily scales to support the largest OLTP throughputs with minimal resource overhead.

The benchmark measured two important aspects of the Informatica Dynamic Data Masking installation:

1. Actual throughputs (in SQL requests per second)

2. Resource consumption (relevant when installed on the database server and not on a dedicated standalone Linux/UNIX machinte)

Test Conditions

This benchmark was conducted on Red Hat Linux 64 bit 16 core Intel Xeon CPU with 32 GB RAM, running with Dynamic Data Masking 9.1.1 on an Oracle 10g database (see Figure 3). To maximize SQL load, multithreaded SQL generating tools were used. The SQL requests included a set of large SELECT requests. Informatica Dynamic Data Masking rules included a set of regular expression matchers to identify SELECT requests with sensitive tables/columns.

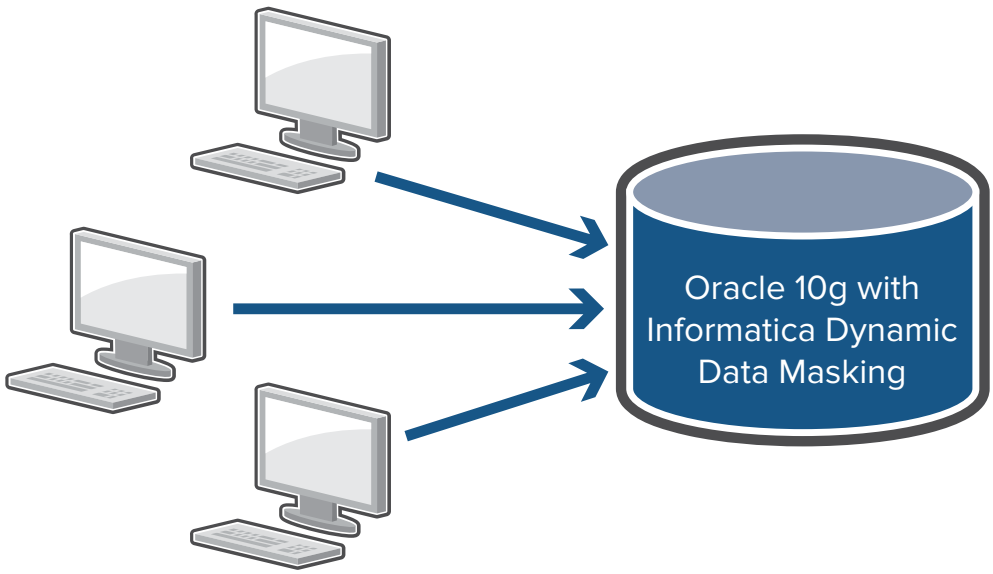


Figure 3: The test methodology of Informatica Dynamic Data Masking used clients with a SQL multithreading generational tool.

Test Methodology

Informatica Dynamic Data Masking was installed on the database server with rules to match sensitive lists of tables and columns. The SQL generating tools, installed on several Windows clients, were configured to first run a stress load to the database server while bypassing Informatica Dynamic Data Masking. The throughput (number of SQL requests per second) and database resource utilization were measured. Three stress runs were then conducted through the Informatica Dynamic Data Masking service, measuring throughput and database server resource consumption.

Benchmark Results

Informatica Dynamic Data Masking responded to SQL requests in 0.15 milliseconds, which is considered negligible compared to the time it would take for the database, application server, and client to process the request. Informatica Dynamic Data Masking is not used for securing batches and ETL processes. It is exclusively configured to secure access by end-users who retrieve information from application screens, reports, and development (or DBA) tools. The benchmark demonstrated that Informatica Dynamic Data Masking supports an application throughput of 3,500 SQL requests per second while consuming only 3 percent of the database server CPU utilization and less than 1 GB of memory.

Informatica Dynamic Data Masking in Action

Informatica Dynamic Data Masking has been demonstrated and proven to protect sensitive information without impacting database performance. This unique advantage has empowered a global mobile communications provider to take a major leap toward preventing unauthorized users from accessing personal data. Prior to implementing Informatica Dynamic Data Masking, this provider had been regularly terminating an average of three people per month for accessing the confidential information of its customers. This process was compromising the company's operational efficiency and damaging its reputation. In an attempt to address the problem, various approaches were explored—yet none of them met the provider's needs for security and performance. Encryption, for example, was ruled out due to performance degradation in the production environment. It would have required numerous changes and continuous updates to applications—a prohibitive task given that many of the applications the organization was using were packaged with closed data models. The company needed a more robust, high-performance approach.

As described earlier in this paper, Informatica Dynamic Data Masking employs a simple visual implementation methodology. This enabled the communications provider to quickly secure a wealth of personal identification data in several of the most complex and demanding business applications, including billing, Siebel, Clarify, and cloned applications. Informatica Dynamic Data Masking allowed personal information to be secured from the company's business users, newly recruited and existing employees, contracted staff, and outsourced and IT staff—allowing them all to access that information while complying with "need-to-know" data access policies.

In addition to dramatically decreasing the risk of a data breach, the software supplied the communications provider with the flexibility to quickly customize data masking capabilities for different regulatory or business requirements. Rule propagation furnished rapid protection across critical production, training, and nonproduction environments. In addition, compliance with privacy regulations was achieved cost-effectively and without any impact to database performance. Furthermore, the company was able to bypass expensive and time-consuming changes to applications that would have incurred lengthy development and testing processes. Commenting on the impact of Informatica Dynamic Data Masking software, the company's chief information security officer said, "In just a few weeks, the Informatica Platform transparently masked personal information on our billing, CRM, and custom application screens and packaged reports in production and nonproduction environments. The Informatica software is now a cornerstone of our risk management and compliance strategy."

Conclusion

Collecting sensitive data is a reality of doing business in the enterprise, including commercial, private, and public sectors. In the process, personal information or company confidential data may inadvertently be exposed to business users, IT staff, developers, consultants, and offshore teams. Organizations must proactively minimize the risk of data breaches by de-identifying data and controlling access to production environments.

Informatica Dynamic Data Masking is high-performance, robust, and scalable data security software that prevents unauthorized users from viewing sensitive information by masking data in real time. It blocks and audits unauthorized access to data, and alerts end users to that access—all while ensuring quick compliance with privacy regulations.

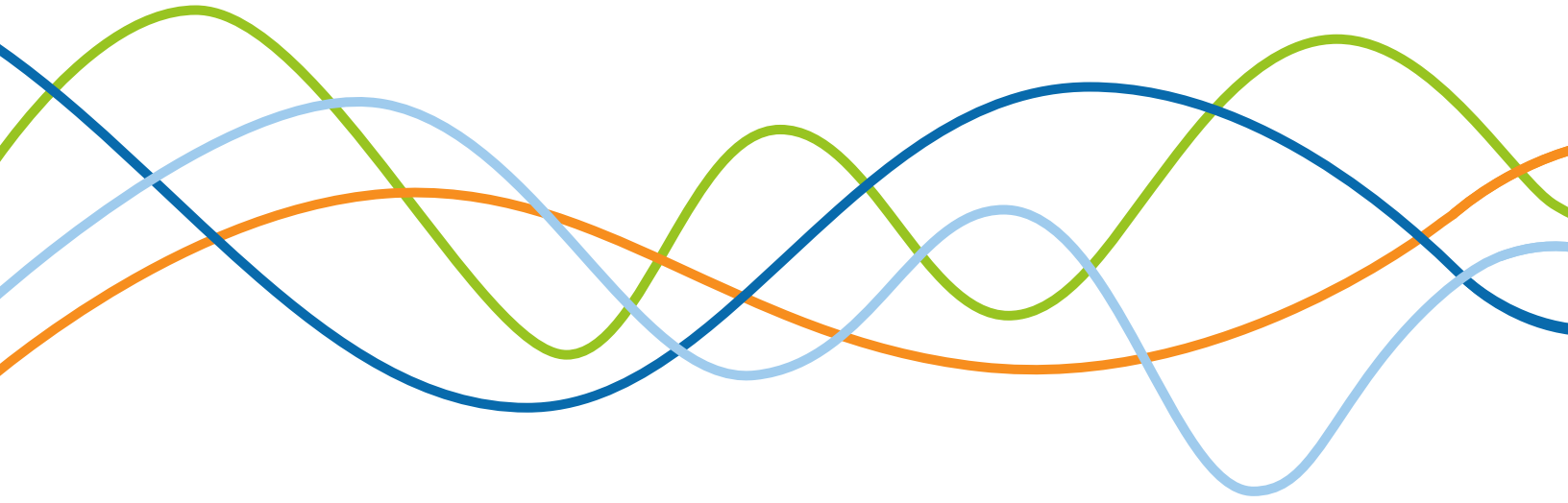
Informatica Dynamic Data Masking offers the following benefits:

- Protection of sensitive information in real time and seamless support for outsourcing initiatives
- Shielding production applications and databases without changes to source code
- High performance by design and architecture substantiated with benchmark results

With its benchmark-proven performance advantages, Informatica Dynamic Data Masking helps organizations to meet the challenges of enforcing data security in the era of big data. Built on an industry-leading data integration platform, the software also helps organizations comply with data privacy policies, regulations, and mandates at lower costs.

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

IN09_1112_02195