INFORMATICA

WHITE PAPER

Secure Sensitive Data in Virtual Test Environments

The Joint Solution of Informatica Dynamic Data Masking and NetApp FlexClone

This edition published February 2013

Table of Contents

Executive Summary

In its 2011 report, The 2011 Digital Universe Study: Extracting Value from the Chaos, research firm IDC estimates that the amount of information enterprise data centers manage will grow by a factor of 50 over the next decade. As data volumes grow, so do the challenges of data security.

While IT organizations are on guard against external threats to data security, internal threats remain an intransigent and common problem. The increasingly complex IT environment makes it difficult to prevent internal data breaches. Most IT organizations need to develop, test, and maintain multiple applications to support individual business units. Realistic data is needed to ensure high-quality development, testing, and training activities. As a result, IT organizations are managing multiple copies of each production application. With each copy made, more users have access to systems containing potentially sensitive or confidential data.

How can IT organizations balance the need to provide development, testing, and training teams with realistic high-quality data in virtual test environments against preventing data breaches? And how can IT organizations protect sensitive data without implementing production controls or requiring the data to be physically masked?

The answer is a joint solution from NetApp and Informatica. Combining the forces of NetApp and Informatica technologies gives IT organizations the best of both worlds: quick clones with a high level of sensitive data protection. The joint solution is based on two products: Informatica® Dynamic Data Masking in concert with NetApp® FlexClone®.

This white paper explores the relationship between growing data volumes and the need for new data security approaches. It examines NetApp and Informatica's different approaches to protecting sensitive data. And it explains the benefits of combining these approaches in a single solution. With this joint solution, IT organizations can:

- Reduce the risks of security breaches
- Boost efficiency in managing their development and test environments
- Lower the total cost of data storage and management

The Relationship Between Data Volumes and Data Security

By virtue of its volume, variety, and velocity, big data is particularly prone to data breaches. 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.[1] As big data volumes grow, new data management strategies need to scale along with them. Virtualization has been one strategy to manage these volumes at the application level. Now, IT organizations are looking to apply virtualization to their databases as well.

But managing exploding data growth is not the only problem—the challenge of data security increases as data volumes grow. Even as IT organizations set up sophisticated barriers to protect themselves from external threats, there often remains the risk of equally dangerous internal threats. Forrester has reported that 70 percent of data breaches are caused by insiders.[2] In a May 2012 Ponemon Institute report, organizations surveyed said 50 percent of data breach cases involve a malicious insider such as a privileged user.[3]

Why is this happening?

The increasingly complex IT environment makes it difficult to prevent data breaches. Most IT organizations need to develop and maintain multiple applications to support individual business units. Developers, testers, and other users require realistic data to ensure all application changes are of the highest quality and rollouts are properly tested prior to promoting them to production.

As a result, for each production application there may be multiple copies—up to 12 full-size copies of production databases for patch, development, test, 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.

[1] IDC, The 2011 Digital Universe Study: Extracting Value from the Chaos, June 2011.
[2] Forrester, Test Data Privacy Is Critical To Meet Compliance, October 2009.
[3] Ponemon, Safeguarding Data in Production & Development: A Survey of IT Practitioners, May 2012.

Different Approaches to Protecting Sensitive Data

Both NetApp and Informatica have created market-leading technologies to prevent data breaches in virtual test environments. Both approaches drive down the overall cost of data yet still fulfill the needs of consumers of that data. Yet these companies take drastically different approaches to providing access to production data for nonproduction usage.

Informatica software creates functionally intact subset copies of production data using a purpose-built application integrated with the time-tested Informatica Platform to extract, transform, and load production data into nonproduction databases. In addition, as production data is moved to nonproduction, all sensitive data (such as credit cards, first names, and national identifiers) is automatically masked utilizing data masking best practices. The end result is a much smaller overall database footprint that reduces the risk of a data breach and is less expensive to maintain.

NetApp technology creates clones of production databases quickly and easily and allows additional virtual copies to be created while consuming a small fraction of the space a traditional full copy occupies. Each virtual copy consumes extremely low amounts of physical storage and gives developers and testers the access to rich production data that they require.

However, when clones are provisioned, all data (including sensitive data) is accessible to end users. The standard approach is to put in place production-like access controls. Production-like access controls ensure access is only given to those users who are authorized to work with the data. But development and testing often require access to a wide swath of application modules.

Another approach is to permanently and irreversibly mask the sensitive data immediately after the initial clone, using Informatica software. This approach to protecting sensitive data works well. However, each change made to the cloned database writes physical records to the database. This means that the cloned database becomes a mix of virtual and physical. If the number of updated records is large (aka, the change rate), the benefit of the virtual copy is diminished.

How can IT organizations protect sensitive data without implementing production controls or requiring the data to be physically masked?

The Joint NetApp and Informatica Solution

Combining the forces of NetApp and Informatica technologies gives IT organizations the best of both worlds: quick clones with a high level of sensitive data protection. The joint NetApp and Informatica solution protects sensitive data without having to install production controls or to physically mask data.

Informatica Dynamic Data Masking makes no changes to the database or calling application. As SQL is issued from Web applications or thick client tools, it is intercepted and the application user is determined (including roles and privileges). If it is determined that the user is not allowed access to the sensitive data, the SQL is slightly altered to ensure that when the sensitive data is returned to the calling application it will either be partially or fully masked (or access blocked completely).

The operation is straightforward utilizing NetApp's SnapManager for Oracle (SMO). Figure 1 shows the high-level architecture of SnapManager for Oracle.

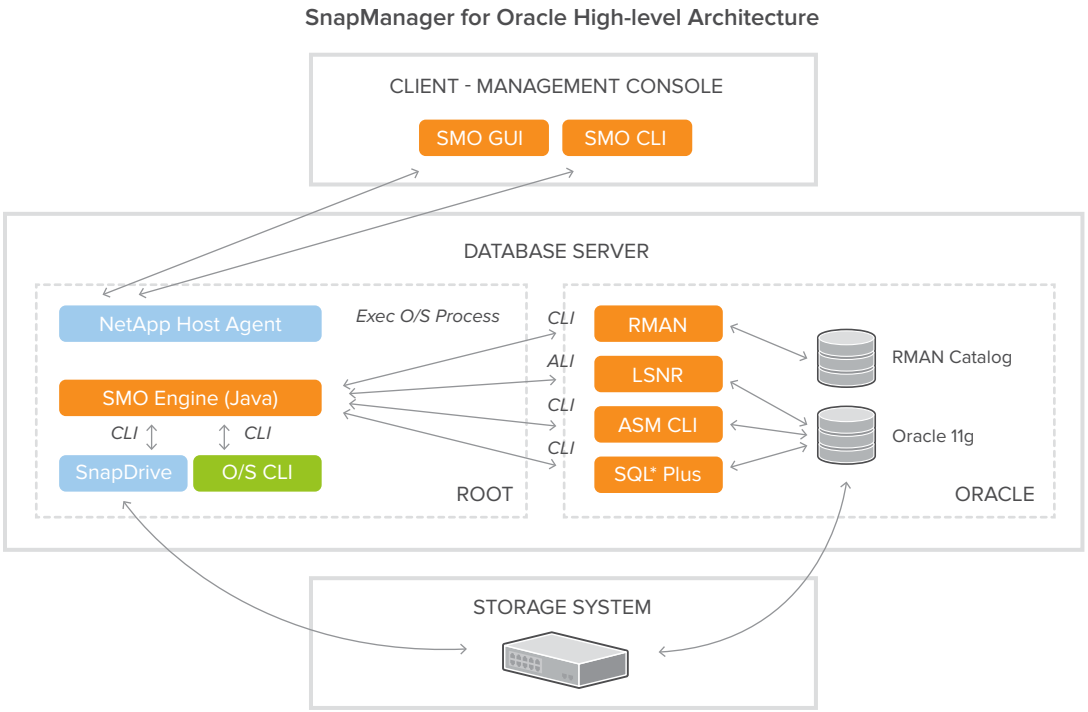


Figure 1: NetApp's SnapManager for Oracle manages the process of protecting sensitive data in virtual test environments.

Once the standard SMO process is completed, Informatica Dynamic Data Masking is installed on the FlexClone volume and the security rules are imported. Informatica Dynamic Data Masking is configured to connect to the database. It rewrites the SQL so that sensitive data, such as phone numbers, Social Security numbers, and credit limits, are blocked, masked, or scrambled. Figure 2 shows how data in a FlexClone is masked in the client for the developer or QA.

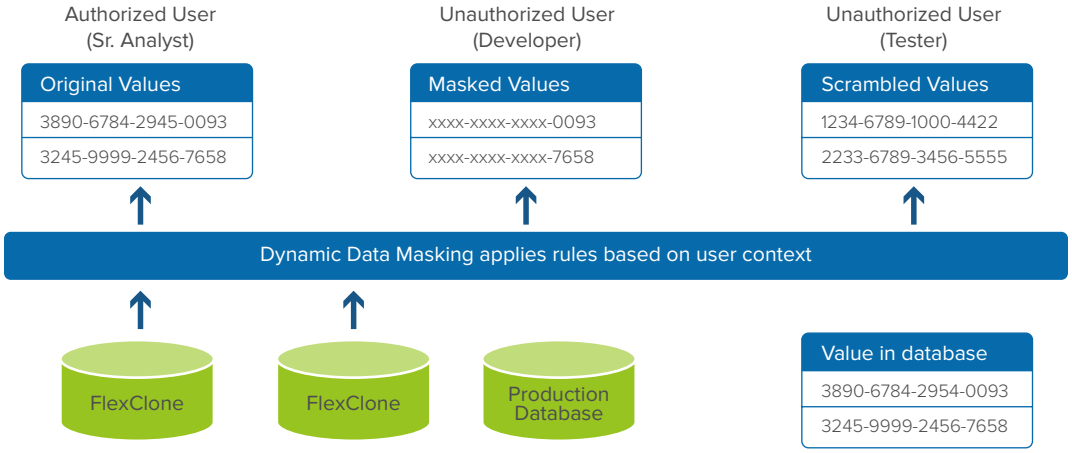


Figure 2: Data in a FlexClone is masked in the client for the developer or QA.

Figure 3 shows an example of the security rules used to mask credit limit and phone number from a customer table.

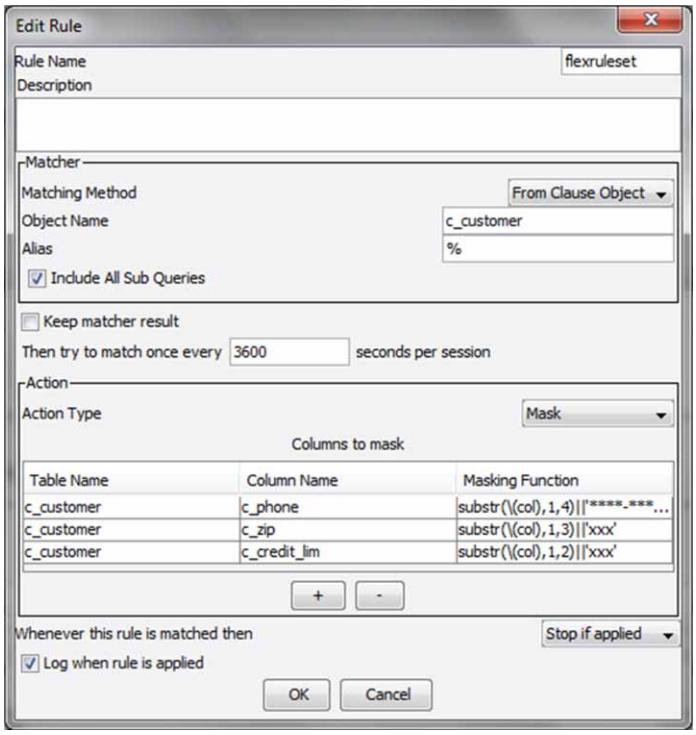


Figure 3: Security rules are used to mask credit limit and phone number from a customer table.

Applying these rules when a user is accessing this customer data masks the data, as seen in the SQL query shown in Figure 4.

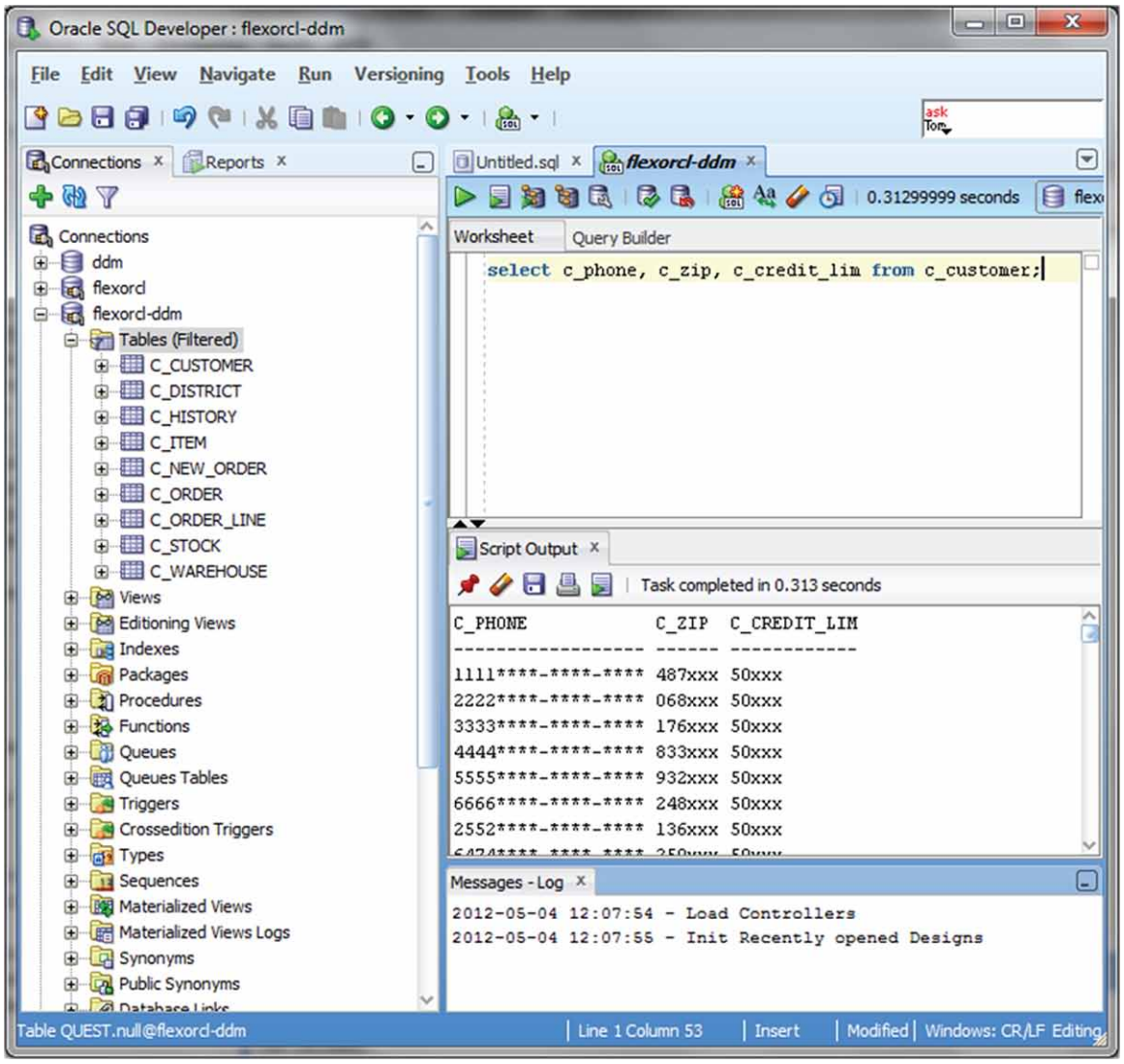


Figure 4: Data is masked as security rules are applied when a user accesses customer data.

Customer Example

A global mobile communications provider was regularly firing 2-4 people a month for accessing confidential customer information. Not only was this costly in terms of human capital management, the data breaches also put the reputation of the whole company at risk. The company needed to stop unauthorized users from accessing sensitive information in production, training, and QA environments.

This solution enabled this telecommunications company to keep personal customer information secure. Business users, newly recruited and existing employees, contracted staff, and outsourced and IT staff have access to this sensitive data on a need-to-know basis only.

The solution's simple visual implementation methodology enabled the IT organization to quickly secure a plethora of personal identification data housed in several of the most complex and demanding business applications, including billing, Siebel, Clarify, and cloned applications.

The solution enabled the IT organization to quickly customize data masking solutions to meet different regulatory or business requirements. Rule propagation furnished rapid protection across critical production, training, and nonproduction environments. As a result, not only does the solution dramatically reduce the risk of data breaches, it also enables the company to comply with privacy regulations and do so cost-effectively. The mobile communications provider was able to bypass expensive and time-consuming changes to applications during its quest for data privacy.

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 solution is now a cornerstone of our risk management and compliance strategy."

Conclusion

Growing data volumes and the increasingly complex IT environment make it difficult to prevent data breaches. IT organizations need to develop, test, and maintain multiple applications to support individual business units. Developers, testers, and other users require realistic data to support high-quality tests and rollouts. They need an easy-to-use, cost-effective way to prevent data breaches in virtual test environments.

The joint NetApp and Informatica solution protects sensitive data without having to install production controls or physically mask data. This joint solution enables IT organizations to:

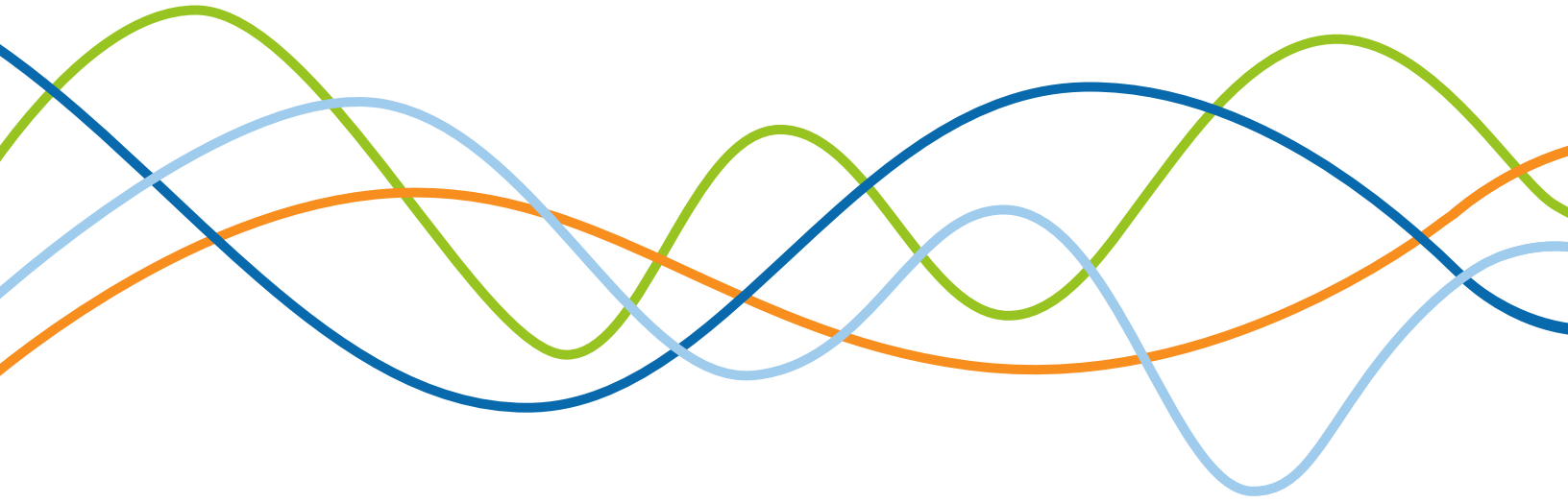
- **Discover sensitive data faster.** With this joint solution, IT organizations can rapidly identify sensitive data across all packaged applications and systems, including legacy apps.
- **Increase dev/test productivity.** Accelerators and prebuilt masking techniques make developing global masking rules fast and easy. Developers can rapidly create multiple thin copies of the data. QA testers can quickly identify optimal test case data and run tests more quickly.
- **Enhance dev/test quality.** Realistic data can be used in development and QA, reducing the need for development rework and production downtime. QA testers can use as many cloned copies as needed to complete QA tests, yielding higher-quality results.
- **Mitigate risk.** With this joint solution, IT organizations can avoid breaches in production data and keep their companies from having to pay victim notification costs or fines. Automated reports on masked data make compliance with data privacy mandates faster and easier. IT organizations can safely outsource application development or support.
- **Lower costs.** This solution boosts IT productivity, shrinks the hardware footprint, enhances compliance, and enables outsourcing—all of which contribute to dramatically lower IT 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.

ABOUT NETAPP

NetApp (NASDAQ: NTAP) creates innovative storage and data management solutions that deliver outstanding cost efficiency and accelerate business breakthroughs. Our commitment to living our core values and consistently being recognized as a great place to work around the world is fundamental to our long-term growth and success as well as the success of our pathway partners and customers. Use of the words "partner" or "partnership" does not imply a legal partnership between NetApp and any other company. Discover our passion for helping companies around the world go further, faster at www.netapp.com.



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_0213_02329