



Why Add Data Masking to Your IBM DB2 Application Environment

Dataguise, Inc.

2201 Walnut Ave., #260

Fremont, CA 94538

(510) 824-1036

www.dataguise.com

Why Add Data Masking To Your IBM DB2 Application Environment

Introduction

Databases are an essential part of any business operation and the data contained in them are critical assets for all entities, big or small. These applications store sensitive data such as credit card numbers, social security numbers, company confidential data and other data, which make them a prime target for data theft attacks. A database is to enterprise data like a bank vault is to money. Databases are critical, and data security is a high priority.

Today database attacks, both internal and external, are on the rise. More than 70 percent of all attacks on databases are from internal sources, making them very difficult to detect and curb. One industry web site, privacyrights.org, reports that more than 500 million customer records have been compromised in the United States since 2005, and these are just the incidents that are known and reported. Large companies ranging from financial institutions, retailers, healthcare providers, insurance companies and prestigious universities regularly suffer negative publicity by reporting data theft or loss. Such attacks are likely to become even more frequent in the future unless organizations take stronger measures to protect their data.

A data security breach can have a severe impact on an organization and consequences can include lawsuits, legal fines, negative brand recognition and decline in stock prices. Managing and mitigating risk is critical to any business, and failure to protect sensitive data from a potential breach can be costly. According to one leading industry analyst firm, the average organizational cost of a data breach incident in 2009 was \$6.75 million.

The Problem: Data is Generally Not Protected in Non-Production Environments

Most enterprises secure production databases when dealing with sensitive data, but only a handful adequately secure their data in non-production environments such as test, development, business analysis and quality assurance (QA). Regardless of where the data is stored, production or non-production, the value of sensitive data to an enterprise remains the same. Unlike confidential paper hard copy, digital information can be duplicated easily and propagated throughout the enterprise, making it particularly vulnerable to disclosure.

The ability to thoroughly test any business application is essential for organizations. In most cases test data comes from production environments, making test databases vulnerable to exposure to both internal privileged users and external hackers. Non-production environments such as test, development, QA and staging databases are often given a lower priority when it comes to security. As a result, managers must deal with the following realities:

Multiple copies of production data generally exist in non-production environments.

On average, five copies of production data exist in non-production environments for every production application instance. These are required to support test, development, QA, migration and staging environments (See Figure 1). More copies of production data means increased security risk.

New user accounts are created without the proper levels of authorization. When a non-production environment is set-up to be accessed by testers, developers, and other IT personnel the new user accounts are typically created with full access to all data. However, the same data in a production environment often has a high level of data access control.

Data is extracted from non-production databases to other, smaller data repositories. Based on requests by customers, vendors and other 3rd party entities, many developers and testers extract company data from non-production environments and output them to files and desktops. This practice not only increases data security challenges, but also makes data vulnerable to unintended exposure.

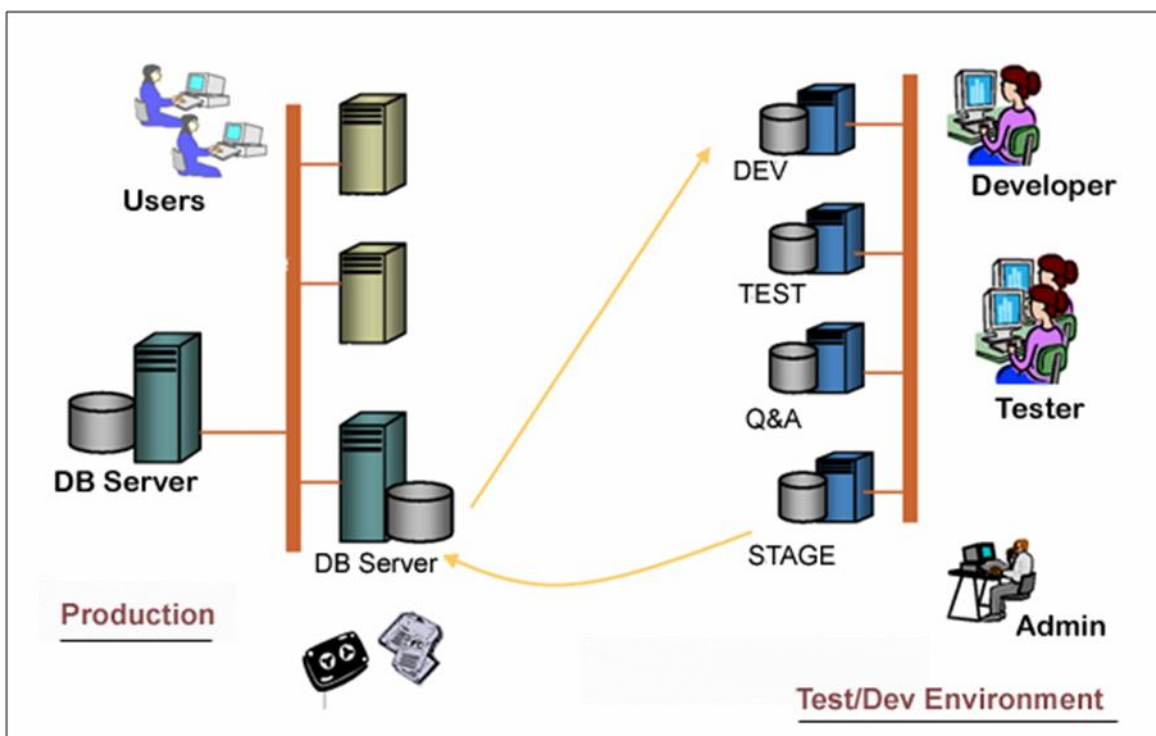


Figure 1: Overview of a typical non-production environment.

Regulatory Compliance Continues To Pressure Enterprises

Over the past few years, various regulatory compliance requirements such as PCI, HIPAA, GLBA and SOX have put pressure on managers. Basic database security such as authentication, authorization, and access control alone are not good enough to meet these various requirements. Organizations now have to implement advanced security measures such as data encryption, auditing, data masking, and real-time monitoring to ensure data privacy and protection. However, each compliance requirement is different. Enterprises need to take the appropriate advanced security measures to ensure they meet these requirements.

PCI Compliance Mandates Strong Data Security Measures

In 2004, Visa and MasterCard inaugurated the Payment Card Industry Data Security Standard (PCI DSS), which is applicable to all major credit-card issuers worldwide. PCI DSS requires that companies establish and maintain adequate internal data security control and procedures pertaining to cardholder information. It covers credit card holder information that is stored or transmitted across the complete technology stack: network, servers, storage, databases, middleware, and applications. PCI directives focus on controls such as strong authentication, access control, auditing, and data encryption. They require establishing strong security policies, processes, and procedures. From a data confidentiality perspective, the top four PCI requirements that require attention from enterprises include:

Requirement 3: Protect stored cardholder data. PCI DSS requires protecting sensitive data wherever it may be production or non-production, off-line or on-line, on-site or off-site, disk, tapes or devices. Recommended approaches to protect databases include data masking for non-production, data-at-rest and data-in-motion encryption for production environments.

Requirement 11: Regularly test security systems and processes. PCI mandates that all enterprises dealing with credit card numbers regularly test their systems from a data privacy point-of-view. Recommended approaches include auditing, monitoring, and encryption.

Requirement 8: Assign a unique ID to each person with computer access. With applications now using web servers and application servers to authenticate users, databases do not have unique IDs to identify users. Therefore, some organizations need to consider integrating application users with database user to keep track of who accesses private data. Recommended approaches include auditing and monitoring of sessions across applications and databases.

Requirement 10: Track and monitor all access to network resources and cardholder data. Enterprises need to ensure that only authorized users can access networks, and applicable controls include monitoring network access and resource utilization. In addition, cardholder data needs to be monitored based on who is accessing or changing such information.

To meet PCI DSS requirements, organizations must take steps to ensure they are protecting cardholder data with controls such as data masking, encryption and monitoring.

HIPAA Mandates All Patient Records Be Protected In All Environments

HIPAA compliance focuses on standardizing communication between health care providers and health insurers and on ensuring the privacy and security of protected health information (PHI). All PHI-related data residing on any database, backup, tape, or transmitted on network needs complete data protection. The key requirements from a database point of view are in Section 164.308 — administrative safeguards — and Section 164.312 — technical safeguards.

To meet HIPAA compliance requirements, enterprises should first ensure that they establish strong authentication, authorization, and access control security measures to applications containing PHI, besides having strong policies and procedures. Enterprises should then look at advanced security measures such as data masking solutions to protect private data in test and development environments. In addition, enterprises should look at data-at-rest and data-in-motion encryption and auditing solutions.

The Solution: Data Masking Helps Protect Non-Production Environments

Data masking, also referred to as de-sensitization, de-identification or data scrubbing, is a process that helps conceal sensitive data. It protects private data in non-production environments such as test, development, and QA, or when data is sent to outsourced or offshore vendors. Data masking changes the original data to de-identify it so that it does not relate to any particular person, entity or context. In the data masking process, the data is changed with special characters such as a hash sign or with new unassociated data.

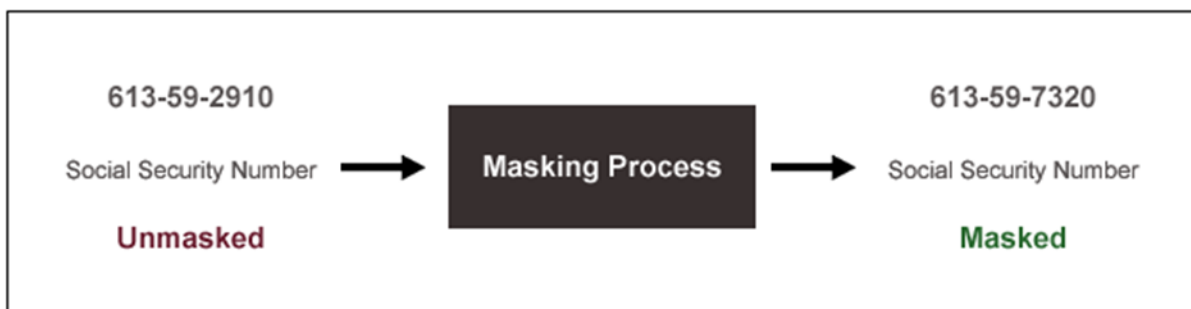


Figure 2: Overview of the masking process.

Data Masking Defined

Data masking is the process of concealing sensitive data so that internal privileged and authorized users of non-production applications cannot access or view actual sensitive data.

A primary requirement for protecting sensitive data with a data masking technology is preserving application and relational integrity with the masked data set. Sensitive data includes social security numbers, bank account numbers, health related personal information, financial information or any company confidential information. Data masking scrambles data to create new, legible data but retains the data properties, such as its width, type and format. Common data masking algorithms include random, substring, concatenation, date aging, sequential, and XOR (bit masking).

Recently the adoption of data masking technology has been driven by a heightened awareness of the need to protect private data in test environments, especially when supporting offshoring or outsourcing of application development. In addition, regulatory and

legal requirements are demanding protection for private data regardless of where it is stored. Forrester estimates that 35% of enterprises will be implementing some form of data masking by 2010, with financial services, healthcare, and government sectors leading the adoption.

What Are the Benefits of Data Masking?

Data masking is the preferred approach for protecting private data, especially in non-production environments such as test, development and QA. The key benefits of data masking include:

Helps meet compliance requirements. PCI and HIPAA mandate that any sensitive data in any database or file be secured and only authorized users have access to such data. Data masking technology helps protect sensitive data in non-production environments such as test and development, by de-identifying data so that privileged users such as testers and developers cannot view it. Testers and developers do not need to view sensitive data for testing or developing an application and such data typically represents only two to three percent of the overall data.

Protects sensitive data in non-production environments. Enterprises often make a copy of production data and use that for testing of applications or for QA purposes. Data masking helps to ensure that any personal or private data is masked before being used in test environments, concealing the data's original value.

Minimizes information risk when outsourcing or offshoring. When outsourcing any application development project or sharing data for any data processing purpose, enterprises should consider masking sensitive data. Most enterprises rely on trust when outsourcing or off shoring data to vendors. Data masking technology improves on this by concealing private data, thus protecting it from being misused or stolen.

What Data To Mask?

Any structured data can be masked. Typically enterprises need to mask only the most sensitive information, which might only be two or three percent of information stored in a database. The goal of protecting such data is to de-identify sensitive information such as credit card or social security numbers. The following are prime candidate data types for masking:

Personal Information. Regulatory compliance requires that customer information should not only be protected from hackers but also be unavailable to privileged users including developers, IT administrators, testers, DBAs and other IT personnel. Consumer- or employee-related information including credit card numbers, social security numbers, and addresses fall under the definition of sensitive data.

Financial Data. All ERP systems contain financial data such as business transactions, profit and loss information, discount information, deal size and revenue information. Such information is not required for developing or testing applications therefore should be masked.

Company confidential. In non-production environments, any other company confidential information should be masked. This might include product roadmaps, employee and executive salary information, and blueprints to technologies.

For any successful data masking project, careful planning is very important along with a deep understanding of the data and how it relates to existing processes. Without data classification initiatives data masking projects are doomed to fail. Proper data protection requires knowing what data exists in the enterprise and where it is located.

Data Masking Algorithms

There are several data masking algorithms that can be applied to protect data in non-production environments. Examples include:

Fictitious data: This technique replaces sensitive data with fictitious values, generating data that looks real when it is in fact bogus. This technique does not typically affect the application or testing requirements because it retains all of the properties of the original data. For example, the customer name "John Barrow" could be substituted with the name "Jim Carlos."

Date aging: In this technique a date field is either increased or decreased, based on policies defined for data masking. However, date ranges must be defined within acceptable boundaries so that the application is not affected. An example would be moving the date of birth back by 2,000 days, which would change the date "12-Jan-1978" to "16-Mar-1972."

Algorithm	Original data	Masked Data	Explanation
Fictitious Data	613-30-3291 (SSNO)	613-30-#### (SSNO)	Last four characters hashed out
Random Data	John Barrow	Jim Arthur	Random data
Date Aging	5/1/2006	3/1/2002	Date decreased by 4 years and 2 months
Numeric Alteration	10201	10401	Numeric increment by 200
Shuffling Data	Jack Mellon	Roger Smith	Name was shuffled

Numeric alteration: In this technique, numeric values are increased or decreased based on a percentage. For example, a salary value could be increased by 10 per cent. This

approach conceals the real value of the data, but if someone knows even one real value it is possible to decipher the entire pattern. As a result, while this is an easy technique to employ, it can also be easily decoded.

Shuffling data: With this approach, data in a particular column is moved to another column or another row. It's like shuffling a pack of cards, and the associations between records and sensitive data are broken. An example would be moving an account number to a random row, so that a customer's account number is different from the original.

Dataguisse Solutions for IBM DB2 Environments

IBM DB2 provides a scalable, heterogeneous platform for enterprise data management. Dataguisse delivers data confidentiality protection for IBM DB2 environments with a solution that locates DB2 instances deployed on enterprise networks, searches DB2 databases to identify and locate sensitive data, and selectively masks the sensitive columns of data in IBM DB2 databases to protect information from disclosure.

Locating and searching databases is accomplished with Dataguisse DgDiscover. In a typical enterprise, virtualization technologies have made it easy for new database instances to be deployed continuously to support software development, QA and test activities. Since sound data security practices require that managers know the locations of all their sensitive data, knowledge of what new or transient database instances are deployed on enterprise networks is essential. DgDiscover can automatically scan specified network IP addresses and ports for active DB2 listener processes. As a result, DB2 instances deployed on the network, including instances deployed in virtual machines, are quickly located and identified.

Once databases are located, the next step is to identify the types and locations of sensitive data in them. Having this information allows managers to understand the different types of data under their care, select the appropriate remediation techniques required to protect the data, and control their sensitive data risks. Dataguisse DgDiscover searches databases to identify and locate columns of sensitive data in DB2 database schemas. This can include PCI data such as credit card numbers and card expiration dates, PII data such as names, telephone numbers, and Social Security Numbers, as well as proprietary data such as account numbers or salary information. DgDiscover returns the types of sensitive data found, their specific table and column locations, as well as a qualitative confidence level that the data is of a certain type.

Masking sensitive columns of data in cloned DB2 databases is accomplished using DgMasker, Dataguisse's enterprise masking solution. DgMasker analyzes DB2 schemas to identify database tables and data types as well as the relationships between columns of data. DgMasker leverages the sensitive data search results from DgDiscover to enable users to directly identify the columns that need to be masked. Users may select an appropriate masking option from a library of pre-defined masking algorithms, or by specifying a custom masking algorithm if desired.

Why Add Data Masking To Your IBM DB2 Application Environment

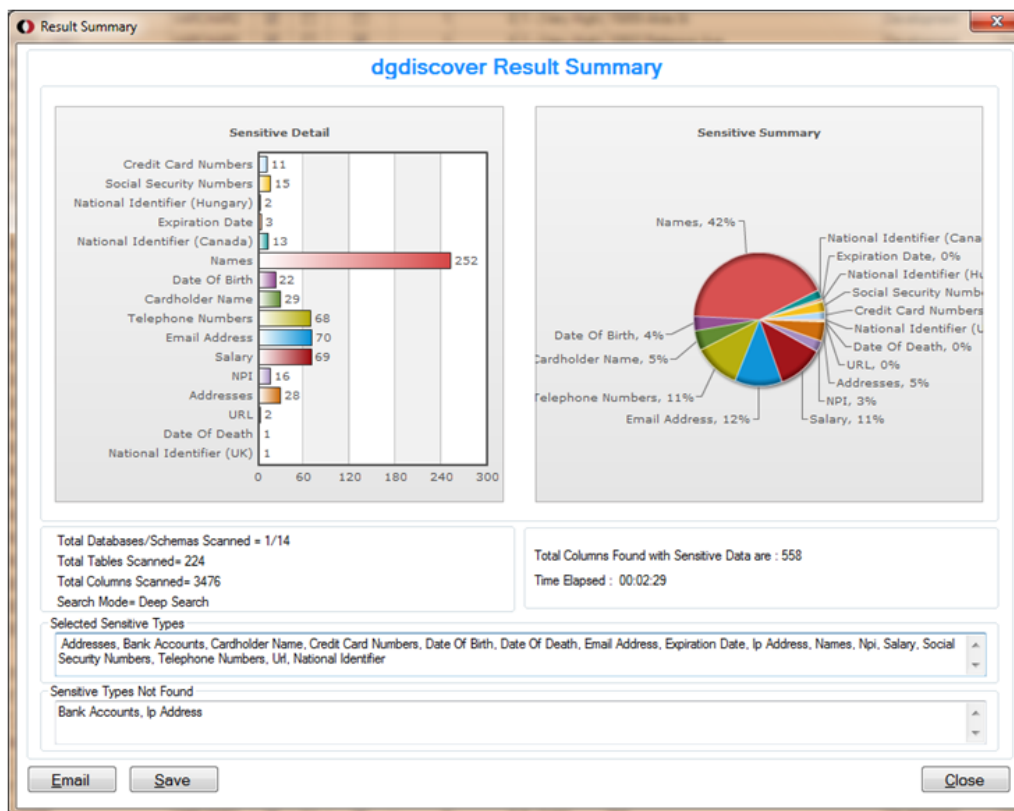


Figure 3: DgDiscover searches DB2 and other databases to identify and locate sensitive data.

In addition to the built-in masking policies available, DgMasker also allows users to specify CUPS constraints for the masked data. CUPS stands for Consistent, Unique, Persistent, and Synchronous. The optional CUPS constraints allow organizations to address specialized masking requirements. For example, the Consistent option can be used to preserve the statistical character of production data with masked data. Unique and Synchronous options allow sites to meet downstream data constraints for application and relational data integrity. The Persistent option lets sites perform data masking the same way across multiple runs to support QA regression testing and the building of accretive data sets in non-production application environments.

Once a masking policy is defined with DgMasker, it may be executed interactively through the application's GUI or via a Java language-based command-line client. The command-line client allows masking to be executed from any platform supporting Java Virtual Machines and enables easy integration with automated software development lifecycle processes. With DgMasker, masking of DB2 tables is performed in-place on the database in DB2's native SQL PL language. This architectural approach to masking allows organizations to fully

Why Add Data Masking To Your IBM DB2 Application Environment

leverage the capacity of their enterprise server platforms to perform masking reliably and efficiently

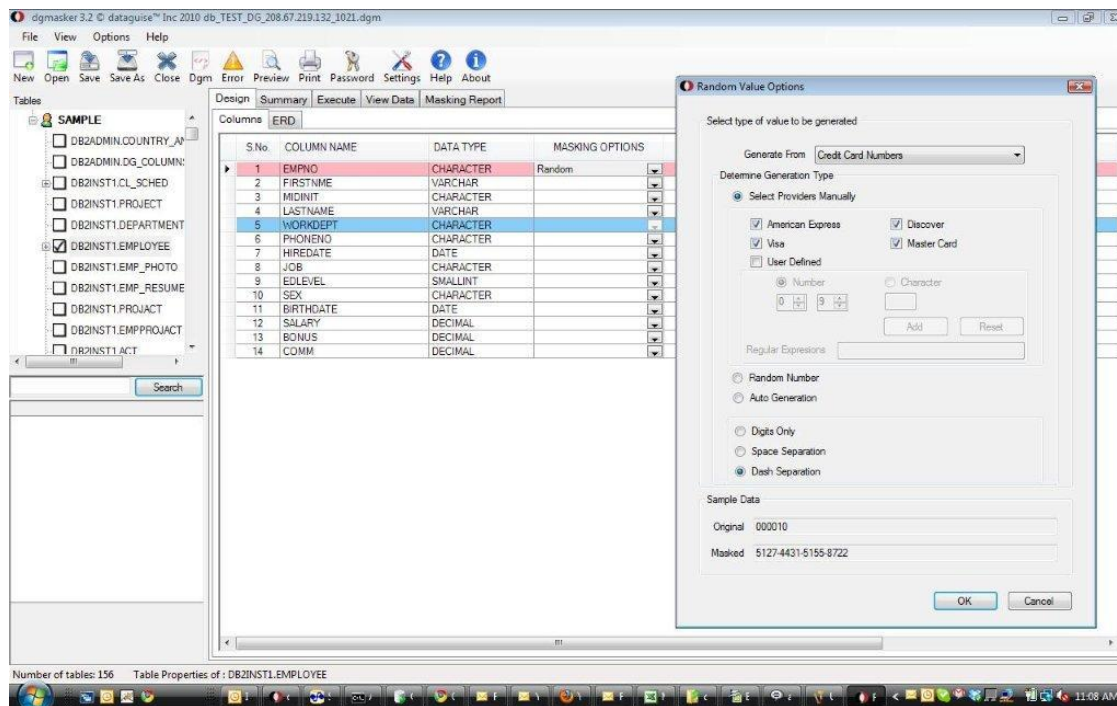


Figure 4: DgMasker facilitates rapid specification of masking policies to enable Masking-On-Demand™.

Dataguis DgDiscover and DgMasker support database discovery, searching, and masking for IBM DB2 environments on Windows, Linux and AIX. Specific DB2 databases supported include versions 9.1, 9.5 and 9.8.

Dataguis Masking Helps Address Compliance Requirements

Dataguis DgDiscover and DgMasker offer a highly automated solution for data discovery and masking which helps enterprises protect sensitive data and meet compliance requirements such as PCI, HIPAA, GLBA and SOX. DgDiscover and DgMasker were built from the ground-up with security and compliance in mind. The solution lets application developers and testers test applications against production data without giving them access to sensitive data such as credit card numbers, social security numbers, account numbers and other data. Key features of DgDiscover and DgMasker include:

- Automated discovery of DB2 and other database instances throughout the enterprise
- Automated identification and location of sensitive data within database schemas

Why Add Data Masking To Your IBM DB2 Application Environment

- Integrated, graphics-based reporting of sensitive data discovery results
- Single, integrated console for defining masking policies for IBM DB2, Oracle and Microsoft SQL Server databases
- Automatic analysis of schema structures and relationships and generation of informative entity relationship diagrams (ERDs)
- Comprehensive library of built-in masking algorithms, with the added ability to support custom procedures quickly and easily
- High performance and scalable masking-in-place architecture to support very large environments
- Preserves application and referential integrity in masked data sets
- Command line client option enabling easy integration into software development processes
- Fast time-to-value with quick deployment and intuitive user interfaces

Conclusion

To meet compliance requirements and prevent data breaches, enterprises need to implement strong protection for all of their private data regardless of where it resides. Data discovery and masking is a superior approach for protecting data when used in non-production environments such as development, test, QA and offline business analysis. For IBM DB2 shops, Dataguise provides a comprehensive solution for protecting sensitive data in these environments with products which automatically locate DB2 instances deployed on the network, identify and locate sensitive data in database tables, and mask the data using next-generation, masking-in-place technology for high performance and scalability.

About Dataguise

Dataguise helps organizations safely leverage their enterprise data with a comprehensive risk-based data protection solution. By automatically locating sensitive data, transparently protecting it with high performance Masking on Demand™, and providing actionable intelligence to managers, Dataguise improves data risk management, operational efficiencies and regulatory compliance costs. Learn more at www.dataguise.com.