Protecting Sensitive Data

A Guide for Enterprises Dealing with A Multi-Shoring Development Environment

Whitepaper

http://www.cogentdatasolutions.com
info@cogentdatasolutions.com

Hoffman Estates, IL • New York, NY • Los Angeles, CA • Austin, TX • Hyderabad, INDIA

COGENT DATA SOLUTIONS

CONTENTS

3 Introduction

3 Facing the Challenge of Organizational Data Security

6 Managing Information Masking Process when Dealing with a Multi-shore Delivery Model

7 Management Solutions for Sensitive Data

10 Core Elements of Data Security Framework

12 Benefits of Data Security Framework

13 An Example of a Project Process Pattern

16 Conclusion

INTRODUCTION

A large number of organizations today are going global, looking for more growth opportunities for their business. However, as these organizations share their data with more and more entities, data privacy and confidentiality is becoming a major issue that needs to be addressed. This paper talks about ways in which businesses that are going global can protect their data in a multi-shoring environment, and a possible pattern proposal is made that can be used by organizations.

FACING THE CHALLENGE OF ORGANIZATIONAL DATA SECURITY

Within a company's data warehouse, a large quantity of sensitive data is stored related to its customers, personnel, finances, healthcare and other business aspects that can be used by a number of different users including DBAs, testers, system developers and application users. From both internal and external sources of an organization, this data is constantly at a threat that can violate the privacy of an organization. Research indicates that most security lapses, in terms of data result, from within an organization and the percentage of these incidents is constantly increasing.

Regardless of whether these security threats result from within an organization or from the outside, the challenge for the organization is to secure sensitive information from these threats. However, there are certain challenges that organizations face when it comes to the security of this data. Some of these challenges include the following:

The Complexity of Data Protection and Privacy

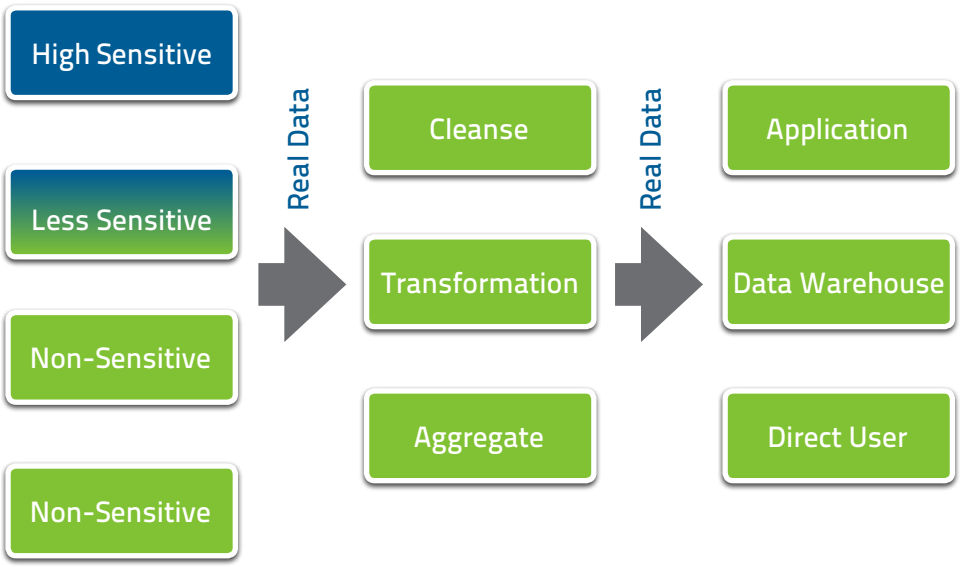
Data protection and privacy measures need to be navigated with business demands and regulatory compliance requirements in consideration and an effective privacy solution should be agile enough to allow businesses to keep up with the dynamic demands of the business environment. The more security measures are taken to protect data, the slower the development pace will get. Therefore, the key issue is to create a balance between data security and the development process by coming up with unique data protection strategies fitting the need of your organization.

The Data Viewing Authority for Organization Personnel

There are some types of information that should only be accessible to related people such as financial and human resource data. Usually, finance and HR managers along with DBAs are the most ideal candidates who should be allowed access to this sensitive information so that they can manage, restore and backup this data. The access to this data should not only be limited to a few people but should also be tracked vehemently. Every time a DBA or authorized personnel access this data, a log should be created so that responsibility can be assigned to authorized individuals for any breach of security. v

It Can Be Difficult To Isolate Sensitive Data

A business operating for a number of years usually has large amounts of data, both structured and unstructured, that needs to be processed and analyzed. Usually all this business data is stored in a data warehouse so that a business can come up with an integrated and time variant data set that can be used for analytical and reporting purposes. BI solutions are also another place where organizational data is stored so that it can be transformed into actionable information. For the implementation of all these structures and systems, an organization would usually have to expose its data to third party developers who can create a threat to their data security.



MANAGING INFORMATION MASKING PROCESS WHEN DEALING WITH A MULTI-SHORE DELIVERY MODEL

When you are dealing with a multi-shore delivery model, strong leadership and business expertise is required by your on-shore and off-shore development teams so that their capabilities can be used at best. They also need to work alongside your clients in order to take a mature approach towards project processing.

At different stages, developers and testers are exposed to the data sources of clients while they are performing functionality and unit tests. Especially in case of BI solutions implementations, developers have to work their way through large amounts of data stored at different locations so that they can process it to create a centralized a data warehouse. All this data should also be made accessible to the off-shore team so that it can be duplicated to a remote network.

To design a successful data privacy program, an end-to-end analysis needs to be undertaken where close examination of problems is conducted through monitoring the implementation procedure. Creating a data privacy program requires organizations to invest heavily in these programs towards activities like training, encryption and attack prevention.

MANAGEMENT SOLUTIONS FOR SENSITIVE DATA

A number of data generation and obfuscation techniques can be undertaken to manage data life cycle, data testing, data masking and enhancing data quality while supporting the existing organizational processes. There are two techniques that can be used to produce realistic data sets in non-production environments so that both time and effort can be saved. The concept behind these two data management components is explained below in detail.

Data Generation and Obfuscation

Data obfuscation is a process where data is scrambled or made more complex in order to protect it from security threats. Unlike data encryption procedures, this data security technique does not slow down processes within your organization that make use of this sensitive information on a regular basis. A number of techniques are followed to transform data into this scrambled structure including:

- Shuffling – In this technique, the order of data is shuffled and the values are moved between rows so that no values are present in their original structure.
- Substitution – Values from a pre-prepared data set are used to play substitute in place of some selected values in the original data, selected in random ways.
- Variance in Number and Date – The data is obfuscated by varying the values in a specified range of options.
- Encryption Algorithm – Using this technique, the data is scrambled into an unrealistic structure so that it cannot be recognized and the volume is data is often increased.

- Deletion or Nulling – With this technique, the more sensitive parts of data are simply nullified or deleted for complete protection.

- Masking – Under this technique, changes made to one part of data are automatically synchronized with other sources that have the same data so that they look realistic.

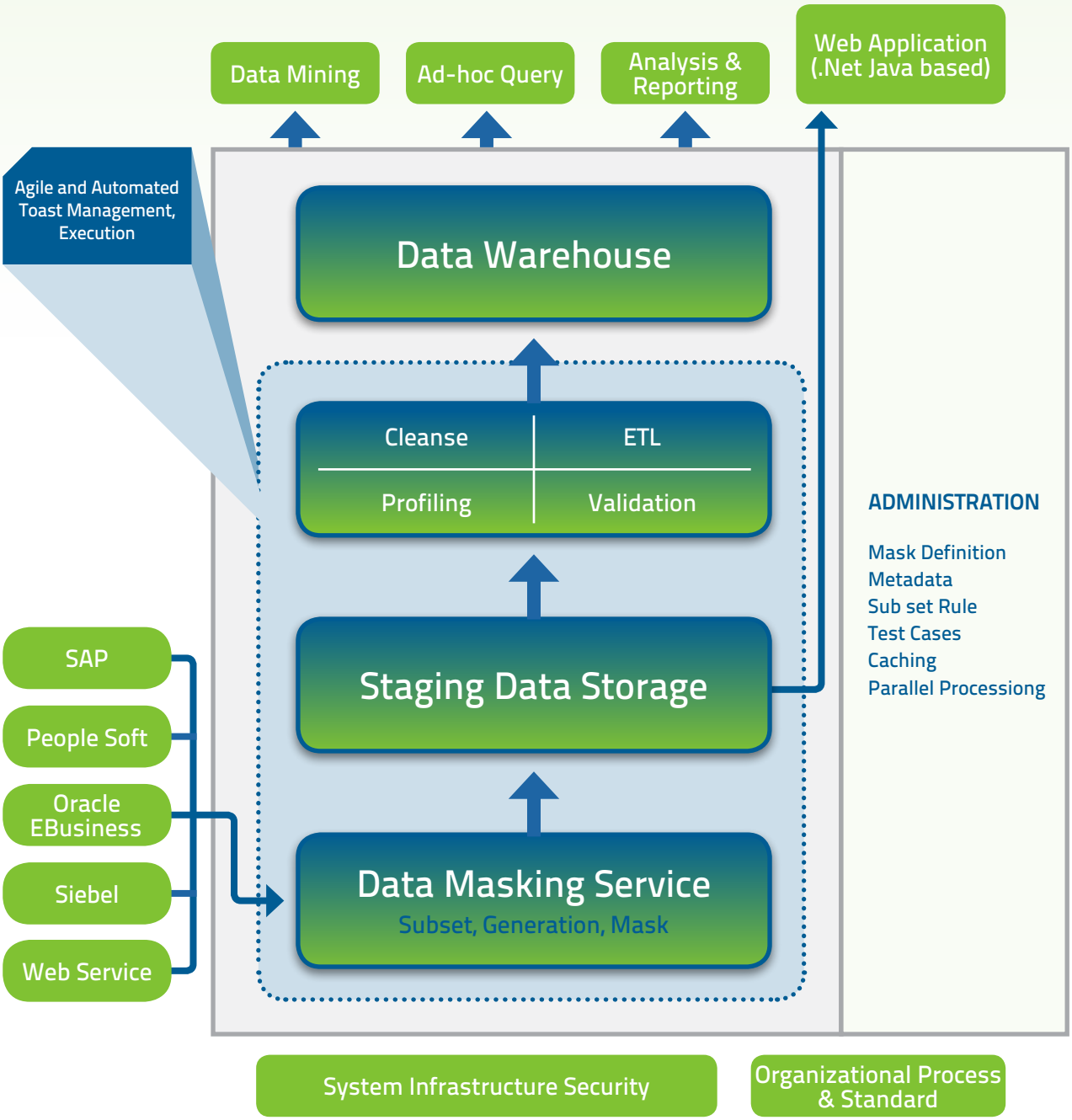
Data generation, on the other hand, is a technique where data testing sets are created with specific rules applicable to each data element. With the use of this technique, realistic business data is created using the above algorithms. Both these strategies can be used in a unified structure to make your sensitive data as secure as possible in multifold ways.

Framework for Data Protection and Management

In order to protect existing data while leveraging its performance and functionality on a daily basis, a data masking approach can be established using different programming languages. The key elements of such an approach may include measures such as:

- Security layers of the system infrastructure can make use of several popular products like Oracle and EMC to create a solid foundation. These elements can be used to create elements like data privacy, password protection and user authentication while complying with the standard business processes to create a proper integration framework.
- An enterprise has various heterogeneous data sources that are used for data extraction by different applications and should all be considered.
- The data masking process is critical and should include all possible techniques so that the data can be made as secure as possible.
- The scrambled data should be made available for further processing depending on its destination. When going to the warehouse, it should remain scrambled while it should assume a real format when it needs to be used for analysis without affecting the performance of BI applications.

Figure 3: Data Protection and Management Framework



CORE ELEMENTS OF DATA SECURITY FRAMEWORK

A number of core technologies can be used with the above framework to make it function at its best capability. These core elements are highlighted here:

Data Sub-setting

This is used when data is being extracted from multiple sources. The volume of data in these sources is extraordinarily large for organizations that have been operating for some years. To process this data, users first need to define selection criteria so that data volume can be minimized. This step is called data sub-setting and users should ensure that the sub-setted data has the integrity of the original production database.

Data Scrambling

While the data is being sub-setted, it can also be scrambled using a defined algorithm on every data column. During the data scrambling process, two points should be considered:

- The original record in the source should only be extracted once for staging and should be reused for the downstream process.
- When multiple data extraction tasks are going on simultaneously, a parallel mechanism should be used for better performance.

Administration

During an application development project, the responsibility of defining a subset, masking algorithm and other such programming decisions lies with the administrator who needs to work together with the QA team to develop an integrated data masking framework.

Extending and Executing Agile Test Management

To reduce time defects and costs, automatic testing procedures are used more commonly. During these data testing procedures, data security needs to be reinforced so that the data masking activities can stay effective.

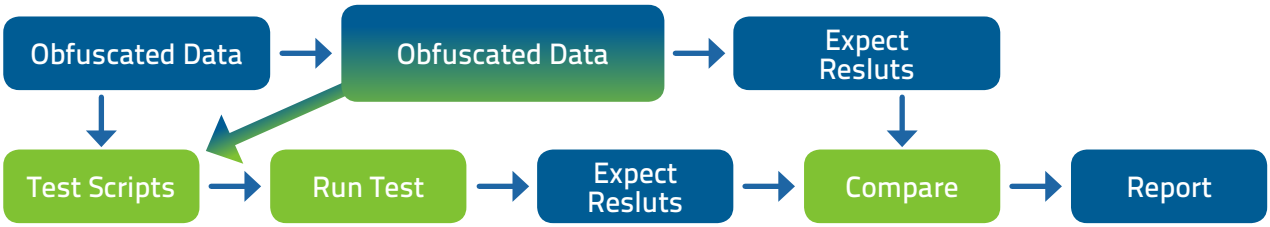
Like all other IT implementation procedures, there are two possible requirements for testing that can be used. The first is the development of new functionality along with the unit test, acceptance etc. while the testers don't have access to sufficient data due to security measures. Second, the use of automated testing is preferable when BI solutions are being implemented or improved.

BENEFITS OF DATA SECURITY FRAMEWORK

With the increasing competition in the security assurance and marketing sector, and the large number of solutions that are available in the market, organizations today are extremely careful about what is more appropriate for their business needs. A preferred data masking approach can bring the following key benefits to an organization:

- Obfuscating Original Information while Maintaining Former Processes The data obfuscation service is treated as an independent element that can be later imbedded into system development and testing processes.
- Creating Realistic Datasets for Testing Using predefined rules, this approach can be used for creating simulated production-like data even in the absence of a production copy.
- Providing Unique and Flexible Data Security Needs Both customized and existing data masking functions can be supported for different types of data so that compliance with corporate data security regulations can be maintained.
- Identifying Data Quality or Defects Using Automated Result Comparisons Coupling automated testing with a data protection framework, identification of potential problems can be made easier and it can be determined how these problems should be addressed for future analysis.
- Creating Seamless Integration with Multi-Shore Teams for Development and Testing After a way of securing sensitive data is established and applied, dividing data information across offshore development and testing teams can be greatly simplified.

Figure 4: Streamlining test steps with obfuscated data



AN EXAMPLE OF A PROJECT PROCESS PATTERN

To make organizations better understand how this data security approach works, a practical process pattern is designed related to the general techniques, activities and best practices.

Initial Context

Before a data protection work stream can be started, some conditions need to be considered including:

- Creation of a team that contain individuals with multiple roles
- Defining project scope, testing scope, and other requirements before the stream is started. This provides clear goals for a data masking process that need to be attained.
- Defining the complete infrastructure including standards, tools and access window so that regular business functions can be carried on unaffected.

Understanding the Role-Based Activities

Within an organization, there may be several other IT projects that are taking place parallel to the data masking project. Roles of individuals along these different activities are often shared. The table below can help you understand four different work stream situations and different elements across them.

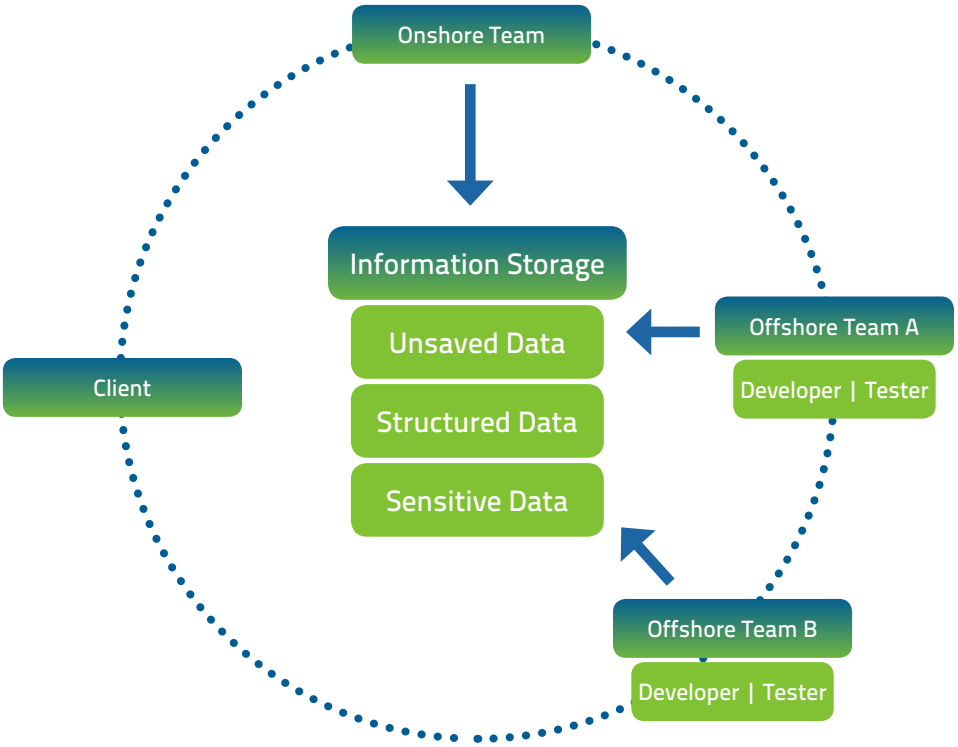
Work Steam	Inputs	Participants	Activities	Outputs
Production Profiling & Analysis	▪ Production Database Schema ▪ Project Requirement ▪ Corporation Privacy Regulation	▪ DBA ▪ Business Analyst ▪ Business Executive	▪ Analyze source schema and related objects ▪ Analyze column prototype (type, length, rules) ▪ Identify sensitivity priority ▪ Analyze data integrity ▪ Gather records statistics	▪ Structured sensitive data scope ▪ Data type & rules on each element ▪ Data quality problems ▪ Data statistics reports
Data Masking & Generation	▪ Data subset tules & scripts ▪ Production database ▪ Project development and testing design	▪ Business Analyst ▪ BI/ Application Developer ▪ Tester	▪ Configure data Masking/ subseting rules ▪ Execute data setting scripts ▪ Execute data masking scripts ▪ Verify	▪ Masked Data sets ▪ Newly generated data sets
Staging Environment Establishment	▪ Production DB schema ▪ Data object identifier	▪ DBA ▪ Business Analyst ▪ Developer	▪ Create new schema ▪ Load transformed data	▪ Staging datasets ▪ Schema description document
Manual or Automated Testing Suite Execution	▪ Project test plan/case/ script ▪ Masked data sets	▪ Tester	▪ Execute test scripts ▪ Generate test report automation	▪ Testing report

Defining Best Practices

The responsibility of identifying sensitive data within an organization and defining data marketing rules lies with the client team. When you are dealing with a multi-shore environment, integration between the on-shore project team and client team is necessary to execute data masking standards while the off-shore team can work on already masked data, finding ways for further development.

Since it is not possible to convert obfuscated data into its original form, this type of data is used as one-off feeds. To identify the quality of obfuscated data, steps are identified to verify that this data meets specific requirements set by data owners.

Maintaining data integrity during the staging process is also necessary so that stage data can match the original data stored in the production database. The data masking functionality should be implemented with agility by prioritizing sensitive data into parts and sub-setting and masking them over a number of weeks.



CONCLUSION

As the business environment gets highly competitive, data security becomes an increasingly important concern for enterprises. However, it is important that businesses find a balanced way to protect their data and share information with offshore teams while maintaining the productivity of their development teams.

The data security and management solutions outlined in this paper are just some of the many practices that are used by organizations today for the protection of sensitive data. A sound data protection and management framework needs to be set in place early on in the project development phase so that proper data security measures are well-integrated with the project development lifecycle.

ABOUT COGENT DATA SOLUTIONS

CDS is a leading firm that provides cutting-edge, innovative business intelligence solutions that ensure that your data is protected and secure. Its services have been implemented across various industries to provide data protection solutions to a wide range of businesses. The success ratio of all these implementations has been extraordinarily high, which is why CDS continues to offer services including BI and beyond.

Known for its cost effective services, CDS always provides an effective and sustainable delivery model for its clients. It offers IT solutions across multiple technology platforms and multiple infrastructures and operating systems. Thriving on seeing positive results for its clients, Cogent Data Solutions has built a superb reputation by delivering comprehensive program management. For companies that need the latest business intelligence tools working in their favor, CDS is their one-stop-shop. Visit http://www.cogentdatasolutions.com for more details.

2500 W Higgins Road
Suite # 1165
Hoffman Estates, IL 60169 USA

Phone : +1 847 238 6262
Fax : +1 866 650 4883
TollFree : +1 866 666 1877
Email : info@cogentdatasolutions.com