Protect your organization's sensitive information and reputation with high-risk data discovery

Locate, identify, and classify sensitive data to reduce data privacy risks, lower potential data sharing exposure, and improve compliance

Table of contents

High-risk data is an important business asset. Are you protecting it wisely?

To operate its day-to-day business successfully, a company must be adept at collecting, storing, processing, and transforming data. These capabilities allow a company to efficiently bill customers, create new products and services, and analyze sales trends to devise targeted marketing plans. This critical information should be protected like any other valuable asset.

As the world has become more reliant on technology, high-tech criminals have also adapted, becoming more and more sophisticated and organized. They can exploit human error and weak security controls to steal trade secrets, payment card data, employee and customer information, and other personal information. Hackers not only rob a company of data, they impugn its integrity, breach its trust with clients and customers, and damage its brand and reputation.

Criminals are not the only concern. Well-meaning employees may lack the tools and training to protect high-risk data. Consider what could happen if someone e-mails an unencrypted file containing high-risk data to the wrong recipient, loses a flash drive containing sensitive information, or leaves a notebook computer unguarded at a restaurant, coffee shop, or other public location. Similarly, organizations may lack the technology, standards, and processes that can help identify, catalog, and control high-risk data throughout the organization.

Awareness of identity theft and personal privacy has never been higher, and employees and customers expect your company to protect their sensitive information. The government and many industry groups require your company to be responsible caretakers of their own and other people's information. Consider the effects of:

- The Health Insurance Portability and Accountability Act (HIPAA)
- The Gramm-Leach-Billey Act (GLBA), which sets rules regarding the protection of personally identifiable information
- The Payment Card Industry Data Security Standard (PCI-DSS)
- The Sarbanes-Oxley Act (SOX)
- Federal and state data privacy laws

High-risk data is any information that, when lost, can lead to significant contractual or legal liabilities; serious damage to your organization's image and reputation; or legal, financial, or business losses. Examples of high-risk information include:

- Name and contact information

  - Name and initials in any combination
  - Home address
  - Home telephone number
  - E-mail address
  - Mobile telephone number
  - Date of birth
- Personal characteristics
  - Age
  - Gender
  - Marital status
  - Nationality
  - Sexual orientation
  - Racial or ethnic origin
  - Religious beliefs
- Personally identifiable information
  - Social security number
  - State-issued identification number
  - Mother's maiden name
  - Driver's license number or similar operating license information
  - Passport number
  - Any other government-issued identification number
  - Credit history
  - Criminal history
- Financial institution data
  - Credit, ATM, and debit card numbers
  - Bank account numbers
  - Financial account numbers
  - Payment card information such as expiration dates, personal identification numbers (PINs), magnetic stripe data, and card verification values (CVVs)

  - Security codes, access codes, and passwords that permit access to financial accounts

- Health and insurance account information
  - Physical and psychological health status and history
  - Disease status and history
  - Medical treatment history
  - Diagnoses by healthcare professionals
  - Prescription information
  - Health insurance information and account number
  - Insurance claim history
  - Medicare and Medicaid information
- Employment information
  - Income
  - Salary
  - Service fees
  - Other compensation information
  - Background check information

Where do you start?

Despite the increasing risks and regulations, some companies may hesitate to take the first step toward implementing a high-risk data discovery or recovery strategy. Some decision-makers will ask what it costs and how it fits into the business strategy. Others may not understand the need review high-risk data concerns, or will not want to address something that may expose risks or that has not presented problems in the past. In other cases, company leaders may not understand their company's exposure to high-risk data loss. They may not know how to articulate their needs, or how to start addressing their challenges.

After a security breach involving high-risk data, companies are required to create an inventory of high-risk data on the affected systems in order to determine what may have been compromised and to inform customers of the breach.

Your company may not know what high risk-data it collects, where the data is stored, or how it is protected. Companies often say they lack these critical capabilities because their information technology (IT) operations grew rapidly to accommodate business requirements, and database administrators simply were not aware of the risk management issues and leading practices involved in collecting and storing certain types of information.

Other companies say that high-risk data became hard to manage because business units developed separate databases to collect and store data, which limits connectivity between IT systems and slows the adoption of enterprise-wide standards and governance.

Introducing processes, technology, and new methods for collecting and storing high-risk data represents an investment of time, people, and money for many companies, and does not readily offer as compelling a return as other IT initiatives.

These concerns are valid, but it is important to remember that high-risk data initiatives, like all exercises in information security, can play a significant role in your IT organization's evolution from asset guardian to strategic business enabler. While high-risk data discovery can be a one-time exercise for your organization, it can also be part of a well-designed, multifaceted security strategy, aligned to business objectives.

Taking the practical approach to high-risk data discovery

Look for a solution that will fit your needs

If time and money were no object, your company could scour every single data byte to discover where its high-risk data is located, and determine the risks associated with each kind of information. Of course, this type of intense scrutiny is rarely necessary or worthwhile. Even if your company had the resources to conduct such a high-risk data discovery initiative, a manual effort would be slow, expensive, and inefficient.

Instead, your company should look for a high-risk data discovery strategy that is practical, efficient, and provides useful and valuable results. When searching for a solution provider, consider the following:

- A solution provider should help you locate the data, determine what kind of information it is, identify its current storage state (that is, whether it is held in the clear, or stored in an obfuscated state such as encryption, truncation, or tokenization), and the risk it may present.
- A solution provider should combine top-down and bottom-up approaches to add specificity to the known high-risk data areas within your systems, while also finding the unknown sensitive data risks. The top-down approach involves talking to people and reviewing processes, while the bottom-up approach uses a variety of tools to search for potentially sensitive information. Combined, these two approaches provide a complete, thorough, and detailed investigation of a company's known and unknown data risks.
- A solution provider should not just rely on technology. Professionals should also work with your company to understand your people, practices, and systems so that they better understand how information is stored in your organization and where high-risk data may be most vulnerable.
- A solution provider should help your company use a wide variety of tools — from leading applications to custom designed programs — to find high-risk data stored in multiple locations as cost effectively, efficiently, and accurately as possible.
- Results from the high-risk data discovery process should help your company address its information vulnerabilities with thorough details, customized reports, data categorization, and risk assessments that can be used to design improvements and remediation action plans.

How PwC helps companies locate, identify, and classify high-risk data

Your company should focus its protection efforts and controls where the need is the greatest: around its high-risk data. To do this, the company must understand the relative value of different classes of information.

PwC can help your company locate, identify, and classify information. PwC's High-Risk Data Discovery team helps companies inventory their data, regardless of physical location, and assign classifications of sensitivity based on legal, compliance, and risk-management criteria.

Organizations store high-risk data in two environments:

	Structured environments	Unstructured environments
What is it?	Ordered databases May be linked to business applications, human resources (HR), or payroll systems, sales and marketing software, or financial applications Information is typically located on servers that are shared either by the entire organization or by individual business units	Information saved outside an arranged database Typically exists in files that can be scattered throughout the enterprise in various documents, saved on servers or individual employees' computers Can also be located on media not connected to the organization's networks, such as thumb drives, optical discs, or portable hard drives Files may be saved and encrypted in myriad formats, depending on customs and preferences of users and work groups
Examples	Oracle databases IBM DB2 Enterprise Server databases Microsoft SQL databases	Microsoft Excel worksheets Microsoft Word documents Text files

Each environment requires a different approach to effectively locate high-risk data.

For structured data environments, PwC has developed a proprietary database analysis tool. PwC's data identification tool scans databases efficiently and precisely, identifying field names or records that might signify high-risk data and then validating actual field values against field format requirements to gain information about the data type, its current state, and its relative risk rating.

For unstructured environments, PwC helps companies take advantage of industry-leading data discovery tools to search for sensitive data in documents that are spread across multiple locations. PwC can also help companies use a data discovery tool to look for sensitive data in e-mail messages and attachments.

PwC takes a thorough approach to data discovery. In addition to analyzing the structured and unstructured data environments, our professionals interview application owners and review documentation to gain a fuller understanding of your company's practices and potential risk areas.

Seeing the big picture — And the smallest details

PwC's Data Discovery Summary Tool provides immediate, interactive access to the findings and results in the following areas:

- **A bird's-eye view of the overall data environment:** All of the servers, databases, and applications in the IT environment, and their relationships to each other
- **Searchable, detailed information:** Detailed information about databases, tables, and fields on each server, along with risk magnitude information for each high-risk data type
- **Custom reports:** Summary reports that can be printed, or exported to Excel

What this means for your business

Your company may be under pressure to locate, identify, and classify high-risk data to meet government and industry requirements. Or you may need to undertake a high-risk data discovery exercise in response to a recent security breach. Or you may decide to pursue high-risk data discovery to gain a more complete picture of your organization's risks and vulnerabilities.

Regardless of the factors that have led you to consider high-risk data discovery services, protecting your company's information is clearly a strategic business imperative. A high-risk data discovery exercise can be part of a broader initiative to protect the confidentiality, integrity, and accessibility of sensitive information around the enterprise.

PwC can help your company design a high-risk data discovery solution that meets your organization's needs and goals. For a deeper conversation about high-risk data discovery and what it can do for your company, please contact:

Laura Guilbert
laura.guilbert@us.pwc.com
415-498-5710

Andrew Toner
andrew.toner@us.pwc.com
646-471-8327