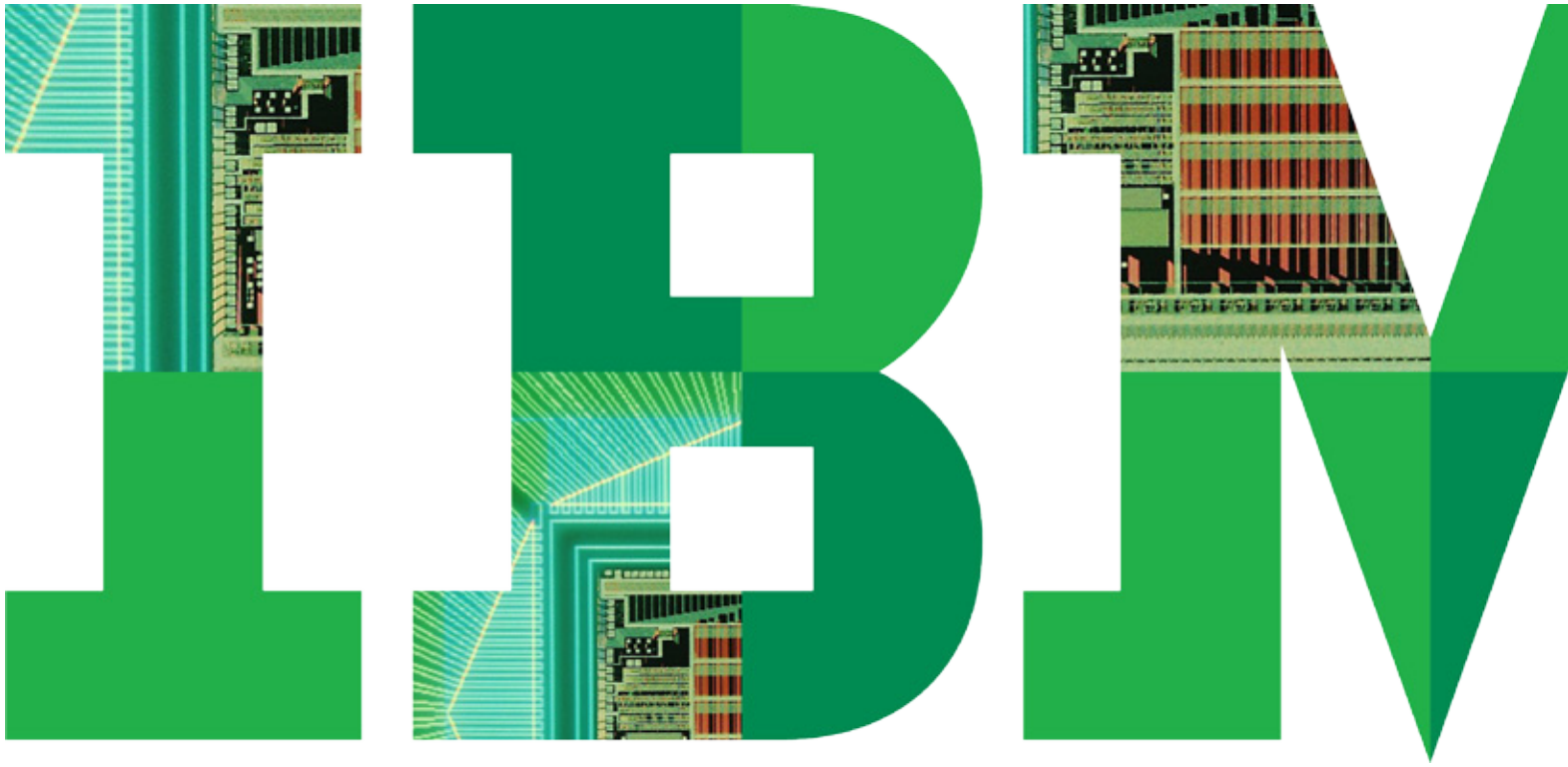
IBM Software

Top Three Myths about Big Data Security

Debunking common misconceptions about big data security



1

The new era of computing has arrived

Harness enterprise information for business advantage.

2

Data: With great value comes great responsibility

Data is powerful and should not be misused.

3

Top three myths of big data security

Understand the true challenges of big data security.

4

Implement big data security

Provide real-time insights and security.

5

Conclusion

IBM InfoSphere Guardium provides security for big data environments

IBM

The new era of computing has arrived

Data volumes have been growing rapidly over the past decade, with no end in sight. Recent research shows that 2.5 quintillion bytes of data are created each day, and 90 percent of the data in the world was created in the last two years. In fact, according to Eric Schmidt, CEO at Google, every two days the world creates as much information as it did from the dawn of civilization up until 2003. Charles Duhigg, New York Times writer and author of The Power of Habit says, "If you are a large company and you are not involved in analytics, you are not a large company for very much longer."

These impressive revelations highlight the opportunity at hand for organizations to capitalize on their own vast information stores. The ability to find insights into new and emerging types of data and content with advanced analytics opens the door to a world of possibilities. Numerous examples of organizations that compete and dominate by implementing a strategy of analytics have emerged during the past few years.

Financial institutions use analytics to differentiate among customers based on credit risk, usage and other characteristics, and then match customer attributes with appropriate product offerings. Harrah's Entertainment hotels, casinos and resorts use analytics in customer loyalty programs. Wineries quantitatively analyze and predict the appeal of their wines. John Deere & Company raised profits 15 percent by employing new analytics tools.

Clearly, organizations need to harness the power of information and analytics to be competitive and gain business advantage. Business leaders are turning to data warehousing platforms and business analytics solutions to generate insight into their operations and gain more value from client opportunities. Using these technologies makes a significant impact, but what about the security of these systems? How can big data platforms be protected? What about the security of the data contained in them? What if the data is misused, or security perimeters are breached?

Security for data warehouses, business intelligence applications and Hadoop-based systems must also be a priority, not an afterthought, especially since they are subject to a growing compliance landscape of internal governance, laws and regulations. Some of the most prevalent mandates include: Sarbanes-Oxley Act (SOX), Health Insurance Portability and Accountability Act (HIPAA), Payment Card Industry Data Security Standard (PCI DSS) (enforcement of which has firmly started expanding beyond North America), Federal Information Security Management Act (FISMA), and the EU Data Privacy Directive.

Deploying big data platforms with large data sets means monitoring transactions while protecting complex data distributed across disparate sources in real time. Data is dynamic. Data moves, flows and changes, and it is combined and manipulated for business analytics.

So how can organizations accelerate business analytics projects while also building centralized, automated security policies for big data systems? The answer lies with a powerful combination of real-time activity monitoring, continuous auditing, business-driven security policies (blocking, terminating, masking, redacting, encrypting), vulnerability management and automated compliance. Establishing a complete access history will allow organizations to understand data and application access patterns, prevent data leakage, enforce data change controls and respond to suspicious activity in real time for big data systems. A secure, centralized repository containing a fine-grained audit trail of all activities across all big data systems eliminates the need to activate native audit functions, and ensures security without high performance overhead.

Security for big data systems is not optional; it's imperative.

Data: With great value comes great responsibility

Data is valuable, and is quickly becoming the new global currency. But data can also be dangerous. The wrong data falling into the wrong hands can have devastating consequences.

Recent examples include: the Sony breach, which cost the company USD170 million in hard costs and potentially more than USD1 billion in lost opportunities, the latest Dark Reading survey reports cybercrime cost consumers about USD110 billion in the last 12 months and affects more than 556 million people, and the annual Deloitte Global Security Study found 1 in 4 financial services firms suffered security breaches in 2011. Add to this the complexity and myriad of regulations and privacy laws, and organizations are stretched to protect all their data and infrastructure across the enterprise.

Hackers are becoming more skilled; they are building sophisticated networks and in some cases are state sponsored. Many organizations are now struggling with the widening gap between hacker capabilities and security defenses. Though often overlooked, a high percentage of data breaches actually emanate from internal weaknesses. From employees, who may misuse payment card numbers and other sensitive information, to those who share privileged access to confidential data, internal controls are critical to monitor insiders. Furthermore, organizations are also accountable for protecting data sent to third parties including business partners and consultants, or other third parties.

According to the **IBM® X-Force 2012 Mid Year** report, “since the last report, we have seen steady growth in SQL injection, which is keeping pace with the increased usage of cross-site scripting and directory traversal commands, such as HTTP ‘DotDot’ commands. These three exploit types become very powerful when they are used together.”

Given today's diverse data sources and the drive to analytics, establishing a process for understanding sensitive data, defining policies, and protecting data (no matter where it resides)—source systems such applications, databases or legacy systems or target systems such as data warehouses, business analytics tools, or Hadoop-based—is challenging.

Common approaches to securing infrastructures include turning on native logging, writing custom scripts to extract and transform data, implementing policies on physical devices, or ignoring security concerns all together. Traditional methods can be labor intensive, error prone, risky and costly, and could require specific skills. Siloed implementations by data source are also extremely risky. Organizations lacking the proper security for their data warehouses or analytics platform increase their risk of a negative event, and could potentially suffer devastating effects such as losing customers, market share, brand equity or revenue.

So, why isn't security of big data systems a top priority, especially given the increased awareness of data breaches and their consequences?

Top three myths of big data security

1) Big data is a buzz word; existing security controls will suffice

Big data isn't entirely new and it isn't just about data volume. Big data is becoming top-of-mind for C-level executives because of the availability of relatively low-cost technology solutions that place big data capabilities within the reach of most organizations, such as Hadoop, MapReduce and Hive.

Big data is a descriptor; it describes a certain category of data. Gartner research has classified big data as data that has not only high volume but also high velocity, variety and complexity. Big data isn't a goal or business initiative. Rather, it is about an organization's ability to process, integrate and understand data to create real-time actionable insight across a secure analytics platform.

New evolving data security strategies require solutions that deliver dynamic protection in real time, are based on business policies, and can scale across emerging platforms. Big data environments are difficult to protect, and present unique challenges:

- Schema-less distributed environments, where data from multiple sources can be joined and aggregated in arbitrary ways, make it challenging to establish access controls.
- The nature of big data comprised of large-scale data sets—high volume, variety and velocity—makes it difficult to ensure data integrity.
- Aggregation of data from across the enterprise means sensitive data is in a repository.

Big data repositories present another data source to secure, and most existing data security and compliance approaches will not scale. According to the **IBM X-Force 2012 Mid Year Trend and Risk Report**, "a more holistic approach to the entire ecosystem is required. Users should become more aware of how visible their personal data is online, more aware of who has access to it, and more aware of how it can be used against them. This affects not only their social networking, but also their choices of mobile application selection and usage. As an increasing trend, mobile applications are requiring a significant amount of permissions that dilute the ability of users to discern potentially malicious intent."

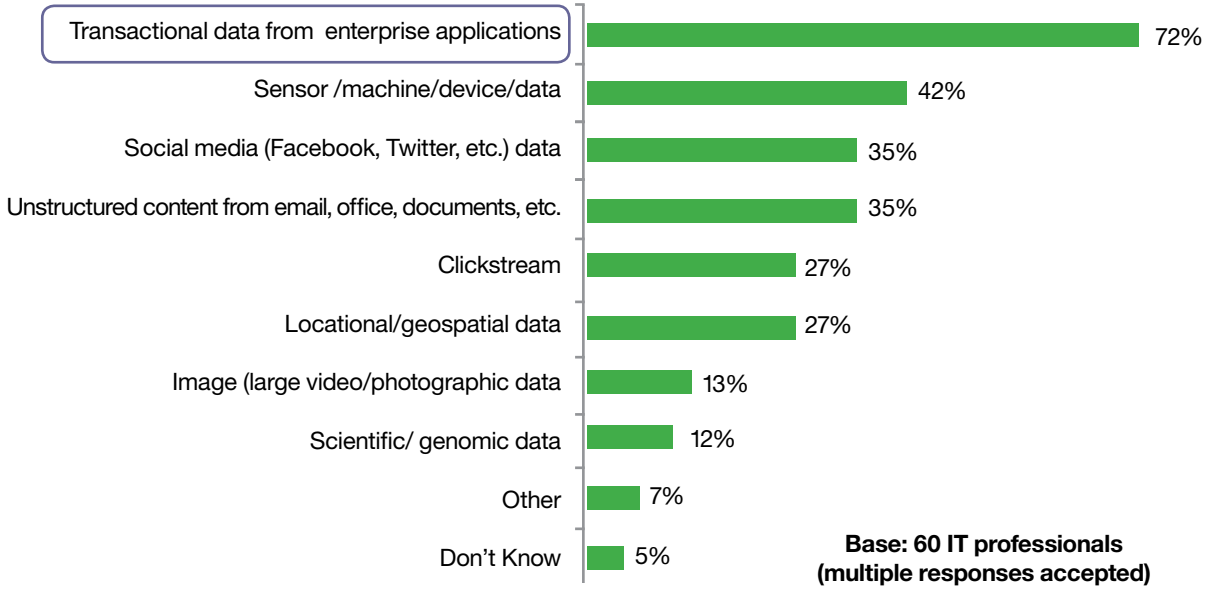
2) Big data deals with social media and sentiment analysis; this data is freely given up by individuals, waiving their security rights

Social media analysis and sentiment analysis are examples of analytics projects. Organizations can be held accountable if this data is misused. Big data projects harness data flowing through organizations at lightening speed in new formats such as social networks, unstructured data repositories, web feeds, sensors, RFID tags, smartphones, videos and GPS data, to name a few. Some other examples of big data use cases include:

- Converting 350 billion annual meter readings to better predict power consumption
- Analyzing 500 million daily call detail records in real-time to predict customer churn faster
- Monitoring hundreds of live video feeds from surveillance cameras to target points of interest

The 2011 Global Big Data Online Survey revealed the number 1 use for big data technologies was to analyze data and transactions from enterprise applications (see Figure 1). This means that databases, like IBM DB2® for z/OS®, Oracle and others, become central to big data initiatives.

“ What types of data/records are you planning to analyze using big data technologies?”



Most big data use cases hype its application for analysis of new, raw data from social media, sensors, and web traffic, but we found that firms are being very practical, with early adopters using it to operate on enterprise data they already have.

Figure 1: 2011 Global Big Data Online Survey, June 2011

Big data is everywhere, not just in social media. In fact, big data repositories like DB2 for z/OS (which contains 60 percent of the world’s financial data), warehouses, and Hadoop-based systems aggregate sensitive data such as PII, business data, HR information and corporate intellectual property. Access to this data should still be monitored and highly protected—even on the world’s most secure computing platforms, such as IBM System z, to guard against internal and external attacks.

3) Big data equals Hadoop and Hadoop is used for internal analytics only

Hadoop is the Apache open-source software framework for working with big data. It was derived from Google technology and put into practice by Yahoo and others. But, big data is too varied and complex for a one-size-fits-all solution.

While Hadoop may have captured the greatest name recognition, it is just one of many classes of technologies well suited to storing and managing high volume, variety and velocity data. Examples of other technologies include IBM Netezza and Teradata.

These systems are managed internally, but how do you verify correct use of data? Plus, many of these high-performance platforms service web requests or other application requests that can expose data and introduce vulnerabilities.

Implement big data security

Big data systems need to be protected against both internal and external threats.

In terms of big data security, there are two distinct areas to consider:

- Security from big data
- Security for big data

"Security from big data" provides the ability to harness data for real-time decision making. When it comes to stopping cyber attacks, real-time data is the difference between safety and disaster. The goal is to deliver risk-prioritized actionable insight. Security intelligence allows organizations to act in real time to stop attacks as they happen.

Examples of big data security projects include:

- Scrutinizing five million trade events created each day to identify potential fraud
- Monitoring hundreds of live video feeds from surveillance cameras to identify security threats
- Catching unauthorized data changes (like log doctoring) as they happen
- Turning on or turning off security policies (like data masking) based on data access patterns

One of the primary drivers of security analytics is the need to identify when an advanced targeted attack (ATA) has bypassed traditional preventative security controls and has penetrated the organization. ATAs are designed to bypass anti-malware scanning systems and intrusion prevention systems (IPSs) and, once established, will attempt to acquire credentialed access, making them extremely difficult to detect.

To be competitive, organizations must improve security decision-making based on prioritized, actionable insight, as well as identify when an advanced targeted attack has bypassed traditional security controls and penetrated the organization.

More detailed user activity monitoring and anomaly detection is required across the entire IT stack. End-user activities should also be monitored and analyzed for anomalous behaviors, including privileged accounts.

"Security for big data" provides protection for big data repositories and the data contained in them.

Security strategies for big data include:

- Sensitive data discovery and classification
- Data access and change controls
- Real-time data activity monitoring and auditing
- Data protection (such as masking or encryption)
- Data loss prevention

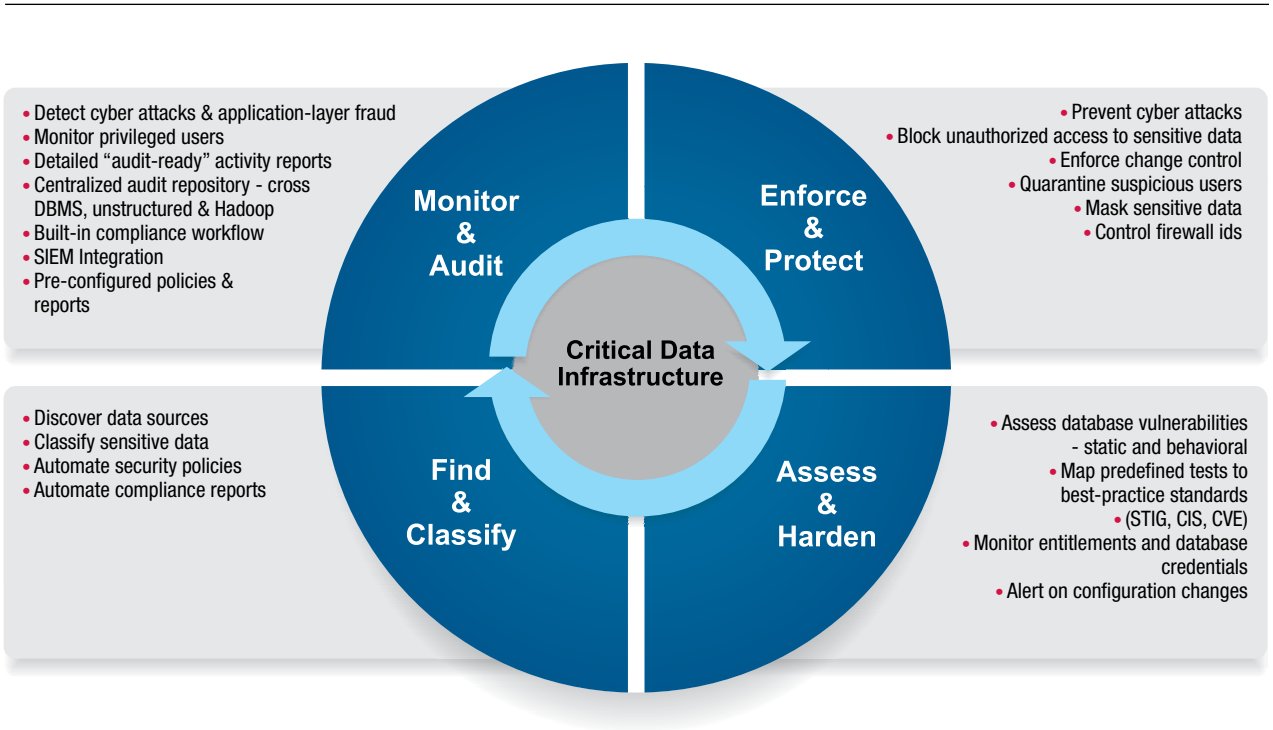


Figure 2: Security for big data is continuous and spans across complex, distributed systems.

- Vulnerability management
- Compliance management

Organizations need to be able to answer questions like:

- Who is running specific big data requests?
- Are users authorized to make requests?
- What analytics requests are users running?
- Are users trying to download sensitive data for legitimate business use or is there reason to question suspicious activity?
- Which transactions are associated with particular application end-users?

Security for big data from InfoSphere Guardium

Effective and efficient data security helps identify security problems quickly and reduce risk in a complex environment. IBM considers security for big data systems a foundational component of any big data project. See Figure 2 for IBM's strategy for big data security.

- Discovery and classification:
  IBM InfoSphere® Guardium® helps organizations create a good mapping of sensitive assets—both of data repositories and sensitive data—inside the databases. Automating the discovery process ensures greater data accuracy and reliability than manual analysis. InfoSphere Guardium also finds malware placed in the data environment.
- Vulnerability and configuration assessment and hardening:
  InfoSphere Guardium verifies secure installation For example, it checks file privileges for database configuration files and executables, and it checks configuration for which privileges have been assigned to critical tables. InfoSphere Guardium integrates with benchmarks such as DISA, STIG, CIS, CVE and SCAP that provide tests to check for common vulnerabilities.

- Change auditing:
  InfoSphere Guardium compares snapshots of secure configurations (at both the operating system level and at the database level) against real-time activity, and immediately and sends an alert whenever a change is made.
- Activity monitoring:
  InfoSphere Guardium provides real-time monitoring of database, data warehouse or Hadoop-based system activity, and collects information from different sources for advanced analytics. Organizations can then create policies based on this security intelligence, such as alerting, masking or even halting malicious activity.
- Auditing and compliance reporting:
  InfoSphere Guardium centralizes reporting across databases, data warehouses, file shares and Hadoop-based systems with a customizable workflow automation solution to generate compliance reports on a scheduled basis. The ability to distribute compliance reports to oversight teams for electronic sign-offs and escalation and to store the results of remediation activities promotes automation and reduces the cost of compliance.
- Data transformation:
  InfoSphere Guardium provides encryption, masking and redaction to render sensitive data unusable, so that an attacker cannot gain unauthorized access to data from outside the data repository. Data transformation techniques protect data in transit; therefore, an attacker cannot eavesdrop at the networking layer and gain access to the data when it is sent to the database client. When data is at rest, an attacker cannot extract the data, even with access to the media files. Employing the correct data transformation technique maintains both structured and unstructured data confidentiality.

Conclusion: InfoSphere Guardium delivers real-time activity monitoring and automated compliance reporting to protect big data environments

InfoSphere Guardium can dramatically simplify your path to audit-readiness by providing targeted, actionable information and data protection. InfoSphere Guardium is the most widely-used solution for preventing data breaches and ensuring the integrity of enterprise data. It is installed in more than 400 customers worldwide and is the first solution to address scalable, enterprise security that protects both data and data repositories, while also automating the entire compliance auditing process.

To learn more about data security visit:
ibm.com/guardium

IBM Corporation
Software Group
Route 100
Somers, NY 10589 USA

Produced in the United States of America
October 2012

NIB03016-USEN-00