Smarter security intelligence

Leverage big data analytics to improve enterprise security

Highlights:

- Integrate security intelligence solutions with big data analytics
- Lower risk, detect fraud and monitor physical and cybersecurity in real time
- Leverage a large volume and variety of data to improve enterprise security
- Act on constantly changing data in motion to prevent potential attacks
- Combat increasingly sophisticated threats, including advanced persistent threats, hacktivism, cyberattacks, physical dangers and insider threats

Big data has been big news in recent years. While organizations have, to date, been busy exploring and experimenting, they are now beginning to focus on using big data technologies to solve business problems. By integrating big data analytics with existing security intelligence solutions, organizations can keep safe, anticipate new attack vectors and act before it's too late.

Five high-value uses for big data

IBM has conducted surveys, studied analysts' findings, talked with more than 300 customers and prospects and implemented hundreds of big data solutions. As a result, it has identified the top five high-value use cases, which could form first steps into big data, as follows:

1. **Big data exploration:** find, visualize and understand big data to improve decision making
2. **360-degree view of the customer:** enhance the existing customer view by incorporating internal and external information sources
3. **Security/intelligence extension:** reduce risk, detect fraud and monitor security in real time
4. **Operations analysis:** analyze a variety of machine data for better business results and operational efficiency
5. **Data warehouse augmentation:** integrate big and traditional data warehouse capabilities to gain new business insights while optimizing the existing warehouse infrastructure

These are not intended to be sequential or prioritized. It doesn't matter where users start; it just matters that they start. The key is to identify which use case makes the most sense for the organization given the challenges it faces today.

This paper focuses specifically on the security/intelligence extension.

The need for stronger security

Security concerns have never been more top of mind for business leaders, consumers and governments. The proliferation of the digital age impacts all aspects of life and is radically changing the way we think about security, in both the cyber and the physical world.

For example, in the largest bank robbery in history, hackers stole USD$45 million without entering a bank, writing a threatening note or using physical force[1]. Instead, a complex web of networks and cyber know-how were the weapons of choice. It took law enforcement officials from 17 different countries to arrest seven individuals.

A recent study revealed that social networks are among the most often used mechanisms to access the defense departments of powerful governments[2]. In fact, during his 2013 State of the Union Address, President Obama cited cybersecurity as one of the top priorities for the United States[3].

Weak password reset rules on 40% of websites allow hackers to penetrate using simple automated programs that run through possible five-letter/number combinations[4].

The digital age also has had implications on the protection of physical assets, such as property, buildings and individuals. Law enforcement agencies must be able to respond to threats by leveraging timely information for a set of predefined behaviors, such as people, vehicles or objects crossing a tripwire or entering a secured area. They must be able to identify and correlate incidents, social media analytics, video surveillance, geospatial records, sensor data — or any other source of data in motion — to proactively identify and monitor potential incidents and create a safer planet.

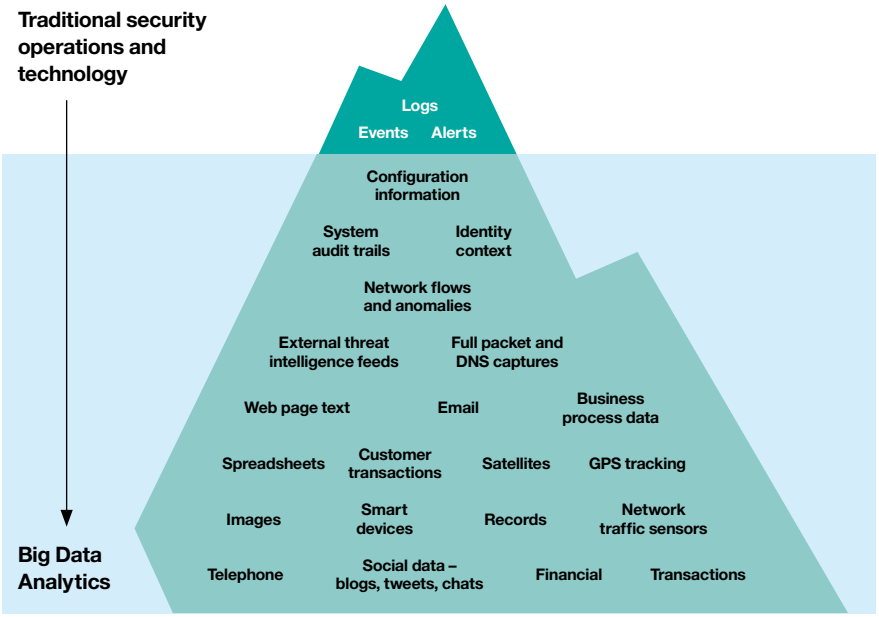
The examples go on and on. Organizations must extend their existing security measures, such as data encryption, monitoring, redaction or masking, to include complementary security intelligence with big data analytics. Security intelligence with big data analytics provides guidance and direction on how to better apply data security and privacy measures across the enterprise.

What big data means for security

We are at a unique point in history: we can now analyze the world around us not only to react to security incidents but also to predict, prevent and take real-time action.

The big data phenomenon presents unique opportunities for organizations to capitalize on data coming from many sources to enhance their security systems (see Figure 1). Sources include traditional structured data as well as new unstructured sources, such as logs, instrumentation data, network data, video surveillance feeds, geospatial information, social data and much more.

More data, in more formats, is speeding across our enterprises. The challenge is to harness it before attackers make their move or identify vulnerabilities.



- Analyze a **variety** of non-traditional and unstructured datasets
- Significantly increase the **volume** of data stored for forensics and historic analysis
- Visualize and query data in new ways
- Integrate with current operations
- Analyze streaming data, keep up with **velocity** of data

Figure 1: Increasingly sophisticated attacks create a need for big data analytics

Security intelligence solutions that make use of big data analytics will help organizations deal with a complex threat landscape. Industry experts recommend innovative thinking and a new approach to security.

"Some organizations will be a target regardless of what they do, but most become a target because of what they do. If your organization is indeed a target of choice, understand as much as you can about what your opponent is likely to do and how far they are willing to go."

Verizon 2013 Data Breach Investigations Report[5]

"Look for security information management (SIM) and network analysis and visibility (NAV) solutions to intersect with big data to enhance security decision-making."

John Kindervag, Forrester Research[6]

"Architect big security data repositories to support security intelligence protection and detection capabilities fulfilled by policy enforcers and scanners."

Gartner Realize That Big Security Data Is Not Big Security Nor Big Intelligence, Joseph Feiman, 19 April 2013

Detecting security threats

Well-organized attackers, criminals and malicious insiders successfully bypass traditional security defenses. The types of attacks they perform are too numerous to explain here, but they can be categorized into six core areas, as listed below:

1. **Advanced, persistent threats:** attackers with both the capability and the intent to persistently and effectively target a specific entity. They have access to a spectrum of techniques and stay low and quiet. The approach is slow, coordinated and adaptable, rather than mindless and automated.
2. **Hacktivism:** hacking for a politically or socially motivated purpose.
3. **Cyberattacks:** attacks intended to damage or destroy a computer network.
4. **Insider threats:** malicious users with in-depth knowledge of the victim's IT enterprise, including privileged accounts, passwords, design and more.
5. **Fraud:** malicious entities attempting to understand and circumvent business process controls for monetary gain.
6. **Physical attacks:** bombs, fires, gunfire or acts of terror on a physical location, such as a stadium, street, home or government building.

As organizations attempt to counter these classes of attacks they need to consider how they "establish a baseline"; essentially, model normal behaviour of applications, databases, users and assets so that anomalies can be early indicators or any of the above forms of attack can be detected.

How to create a security intelligence platform with big data analytics

There are four requirements for a security platform in the new era of computing.

1. Monitor network behaviors to detect known and unknown cyberthreats

The first core requirement is to protect networks from attacks. Proper protection requires the ability to identify advanced threats, such as "botnets." A botnet is a network of compromised computers controlled by a "botmaster." It infects organizations through malicious websites, emailed executables and contaminated memory sticks or PDF files. Botnets range in size from hundreds to millions of hosts and generate a significant amount of traffic on the network. However, the malicious traffic is often overlooked and not understood because transmission control protocol/internet protocol (TCP/IP) volume is extremely high and contains huge amounts of domain name system (DNS) traffic.

In order to protect networks and identify botnets, organizations must monitor all activity on a TCP/IP network to pinpoint anomalous activity that may present a threat. They can do this by leveraging analytical algorithms to find subtle threat indicators.

Big data analytics are critical because they can detect, identify and monitor a bot by analyzing a large amount of data in real time. Solutions monitor all traffic, such as DNS requests, black/white lists and database access, to understand the threat. This analysis must be integrated with external intelligence to develop a complete threat profile.

2. Detect data leakage

To understand if data is being leaked, organizations first need to monitor the information that employees or bots see and/or access. Next, they must determine whether employees are accessing sensitive information — even if they have the right to access it — and figure out if the access patterns are suddenly changing. A customer service representative normally accesses client records, but if the average view per day is 25 and an employee suddenly views 500 records, there is likely cause for concern. The organization must determine if and what confidential or sensitive information might be at risk.

Understanding data access and change history not only makes good security sense but also is required by many common industry and government mandates, such as PCI DSS, HIPAA and SOX.

Big data analytics will improve surveillance by leveraging new data sources and types, such as internet, satellite, video and audio, to make broader correlations and pattern matching possible.

3. Incorporate forensic, fraud and criminal intelligence

The third requirement is to track criminals in real time to predict and prevent crime. This means monitoring various open and covert sources of information to determine the following:

- Who is talking to whom about what and how?
- What do people think about a certain person, organization or government?
- What are the activities, locations, interests, plans of persons and groups, especially of those on a blacklist?
- Is there any other suspicious content?

Examples of data that need to be analyzed include database activity patterns, call data records, web traffic and physical sensors. Advanced analytics on social networks, geospatial information, text, image, video and voice data is required to identify and predict behaviors to prevent crime.

4. Support and enhance law enforcement

Real-time data-mining analytics on the location parameters of street gangs or persons of interest can be obtained through geospatial location detectors, such as GPS-enabled cellphones, and can help authorities to predict and proactively prevent criminal activities. Smarter surveillance includes powerful predictive analytics on multiple concurrent streams of structured and unstructured surveillance data, which can be drawn from such sources as manned and unmanned vehicles and security cameras in real time to alert law enforcement agencies of potential security issues. Blended with real-time name recognition and identity insight, this approach can help law enforcers gain insight into cross-border and intra-border transactions.

Four core capabilities must come together to deliver these three requirements: a security intelligence platform, a streaming data analytics platform and a Hadoop-based analytics framework. This is illustrated in Figure 2.

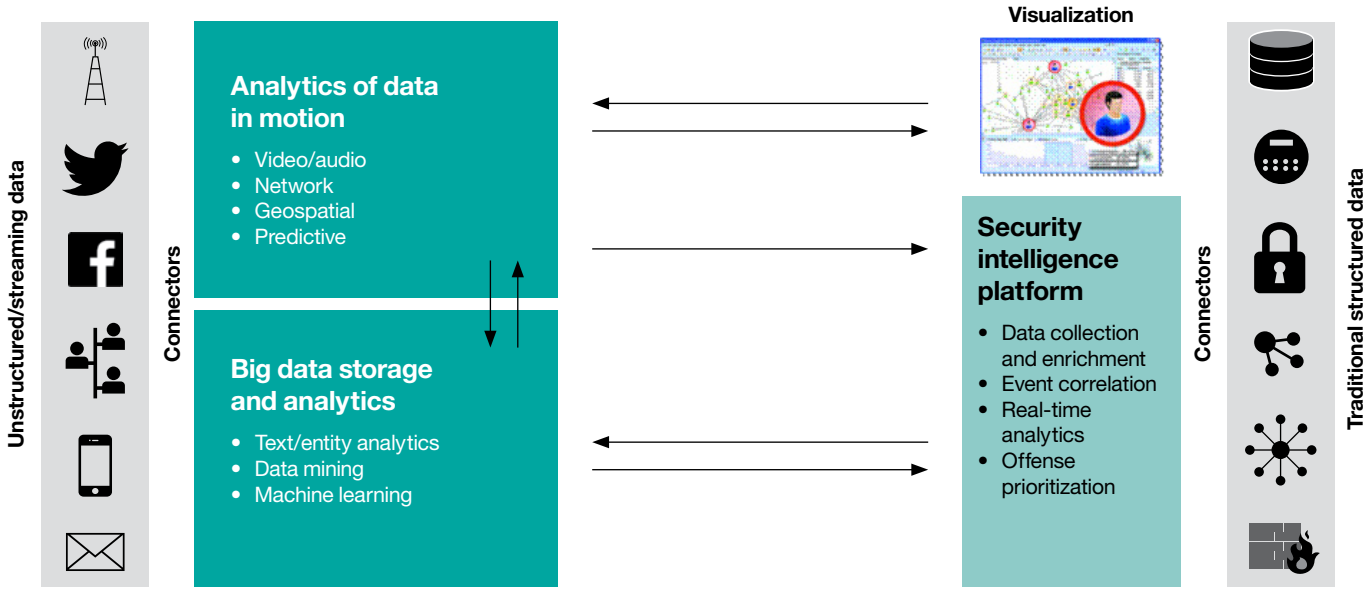


Figure 2: Security intelligence with big data

IBM Quick Start for big data technologies

Free to download, quick to start. To help clients get started with big data, IBM has launched the Quick Start program, which gives users fast access to market-leading big data technology. Available through a free download, the market leading, enterprise-grade big data platform empowers users to experiment with big data in their unique, non-production environment. There is no data or time limit.

- **IBM InfoSphere BigInsights Quick Start Edition –** www.ibm.com/software/data/infosphere/biginsights/quick-start/
- **IBM InfoSphere Streams Quick Start Edition –** www.ibm.com/software/data/infosphere/streams/quick-start/

Benefits of security intelligence with big data analytics

When big data analytics are added to security intelligence, the benefits are straightforward yet significant. First, it lowers risk, helping users to understand potential threats before they happen and to act in real time. Second, it helps detect fraud, identifying baseline behaviors and immediately highlighting anomalies when they occur. Third, it monitors the organization's security status at all times. This intelligence allows users to react in order to prevent crime.

Overall, security intelligence with big data analytics helps organizations take advantage of the large volumes and wide variety of data speeding across their enterprise in order to anticipate and predict threats and to act on predictive analytics. It allows organizations to analyze constantly changing data in motion and perform sophisticated analytics on captured data.

Many IBM clients are taking advantage of the opportunities and seeing measurable results. For example, one US security facility analyzes 1.6GB/second of video data in real time. Another client, a defense agency, identified 43 infected hosts across hundreds of possible domain names. The client detected more than 350 documents with the keyword "confidential" being transferred over the network and more than 900 Facebook accesses within a 45-minute window.

US high-security facility processes 1.6GB of streaming data per second

IBM business partner TerraEchos, Inc., a leading provider of covert intelligence and surveillance sensor systems, provides organizations with advanced security solutions for critical infrastructures and extended borders. One TerraEchos client is a national science-based, applied engineering laboratory dedicated to supporting the mission of the US Department of Energy (DOE) in nuclear and energy research, science and national defense.

The lab turned to TerraEchos to implement an advanced covert security and surveillance system, based on its TerraEchos Adelos® S4 System, an IBM System x3650 server and IBM InfoSphere® Streams software. The solution offers advanced fiber-optic acoustic sensor technology licensed from the US Navy.

Serving as the underlying analytics platform, InfoSphere Streams software enables the Adelos S4 solution to analyze and classify streaming acoustic data in real time. InfoSphere Streams software collects data from multiple sensor types and enables associated streams of structured and unstructured data to be incorporated into an integrated intelligence system for threat detection, classification, correlation, prediction and communication by means of a service-oriented architecture.

Because the solution captures and transmits real-time, streaming acoustical data from around the lab premises, security staff members have unprecedented insight into any event. The system enables lab and security personnel to "hear" what is going on, even when the disturbance is miles away. They can confidently identify and classify a potential security threat and take appropriate action. Intrusions are classified as biological, mechanical or environmental and plotted in a spatial domain, helping to differentiate an animal from a human or mechanical intruder.

"The US Government has been working with IBM Research since 2003 on a radical new approach to data analysis that enables high-speed, scalable and complex analytics of heterogeneous data streams in motion. The project has been so successful that the US government will deploy additional installations to enable other agencies to achieve greater success in various future projects."

US government

IBM big data analytics portfolio

IBM is uniquely qualified to support security intelligence with big data analytics, as illustrated in Figure 2.

IBM QRadar® Security Intelligence Platform consolidates log source events and network flow data from thousands of devices, endpoints and applications distributed throughout a network. It performs immediate normalization and correlation activities on raw data to distinguish real threats from false positives. As an option, this software incorporates the IBM Security X-Force® Threat Intelligence application, which supplies a list of potentially malicious IP addresses, including malware hosts, spam sources and other threats. And the QRadar Security Intelligence Platform can correlate system vulnerabilities with event and network data, helping to prioritize security incidents.

IBM InfoSphere BigInsights™ software builds on the Apache™ Hadoop® framework to include analytics and makes it easier to identify security threats. IBM does not fork the code and maintains Hadoop API compatibility. IBM packages the key open-source Hadoop components in InfoSphere BigInsights software but also includes IBM technology that is not available with open-source Hadoop distributions.

InfoSphere BigInsights software can help improve the accuracy of analysis over time and feed insights to the QRadar Security Intelligence Platform, providing a facility for closed-loop, continuous learning. The result is an intelligent, integrated solution that helps collect, monitor, analyze and report on security and enterprise data in a manner not previously possible.

IBM InfoSphere Streams software is an advanced computing platform that allows user-developed applications to quickly ingest, analyze and correlate information as it arrives from thousands of real-time sources. It can handle very high data throughput rates, up to millions of events or messages per second. InfoSphere Streams software is designed to analyze data in motion, and it provides sub-millisecond response times, allowing users to view information and events as they unfold: a must for security intelligence. InfoSphere Streams software can help improve the accuracy of analysis over time and feed insights to the QRadar Security Intelligence Platform, providing a facility for closed-loop, continuous learning.

InfoSphere Streams software offers a simple development environment. Users can build their own applications using an Eclipse-based integrated development environment with drag-and-drop capabilities and visualization. InfoSphere Streams software delivers a scalable architecture to integrate both structured and unstructured data sources and connect to other security technologies.

IBM PureData™ for Analytics appliance, powered by IBM Netezza technology, is a simple, smart data appliance for serious analytics. It simplifies and optimizes performance of data services for analytic applications, enabling very complex algorithms to run in minutes instead of hours, delivering speed (10 - 100 times faster than traditional custom systems), faster time to value (5TB/hour load speed) and simplicity.

All the above offerings are integrated with IBM's data security and privacy portfolio, featuring **IBM InfoSphere Guardium® and IBM InfoSphere Optim™**[7]. InfoSphere Guardium Data Activity Monitor provides centralized controls for real-time data security and monitoring, fine-grained database auditing, automated compliance reporting, data-level access control, database vulnerability management and autodiscovery of sensitive data.

IBM InfoSphere Optim delivers both static and dynamic masking on demand. Security intelligence with big data analytics helps organizations to intelligently apply security and privacy policies across the enterprise.

Do you need a security platform equipped with big data analytics?

If you find yourself answering "yes" to the following questions, you may need a security solution that includes big data analytics sooner rather than later:

- Do you want to analyze and correlate broader data sets to prevent cyberattacks, physical threats, fraudulent claims or account takeovers?
- Do you need to enrich your security solution with email, social and other unstructured data to improve cyberthreat detection and remediation?
- Do you need to better detect and monitor criminal and terrorist activity by correlating a broader variety of sources to uncover associations or patterns?
- Do you want to enhance your security and surveillance systems with real-time data from video, acoustic, thermal or other devices/sensors?

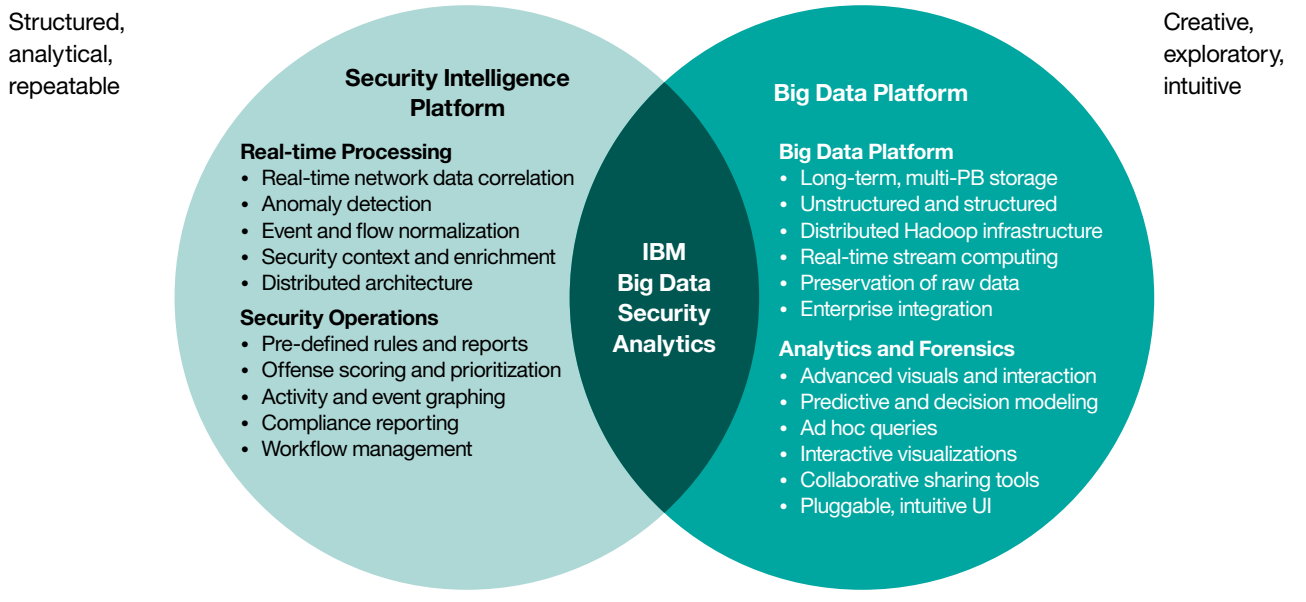


Figure 3: Integrate security intelligence and big data analytics

If you would like to explore the security/intelligence use case further, contact your local IBM sales representative, or visit www.ibm.com/software/data/bigdata/use-cases.html

1 http://thehackernews.com/2013/05/the-biggest-bank-robbery-in-history.html
2 http://thehackernews.com/2013/05/us-department-of-defense-officials-are.html
3 www.bizjournals.com/washington/blog/fedbiz_daily/2013/02/obama-confirms-cybersecurity-order-in.html?page=all
4 http://thehackernews.com/2013/08/short-password-reset-code-vulnerability.html
5 www.verizonenterprise.com/DBIR/2013/
6 "Control And Protect Sensitive Information In The Era Of Big Data", Forrester Research, Inc., July 12, 2012.
7 www.ibm.com/software/data/security-privacy/

© Copyright IBM Corporation 2013

IBM
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America
October 2013

All Rights Reserved

IBM, the IBM logo, ibm.com, BigInsights, Guardium, InfoSphere, Optim, PureData, QRadar and X-Force are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at: **ibm.com**/legal/copytrade.shtml

Other product, company or service names may be trademarks or service marks of others.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates.

Please Recycle

IMS14435-USEN-00