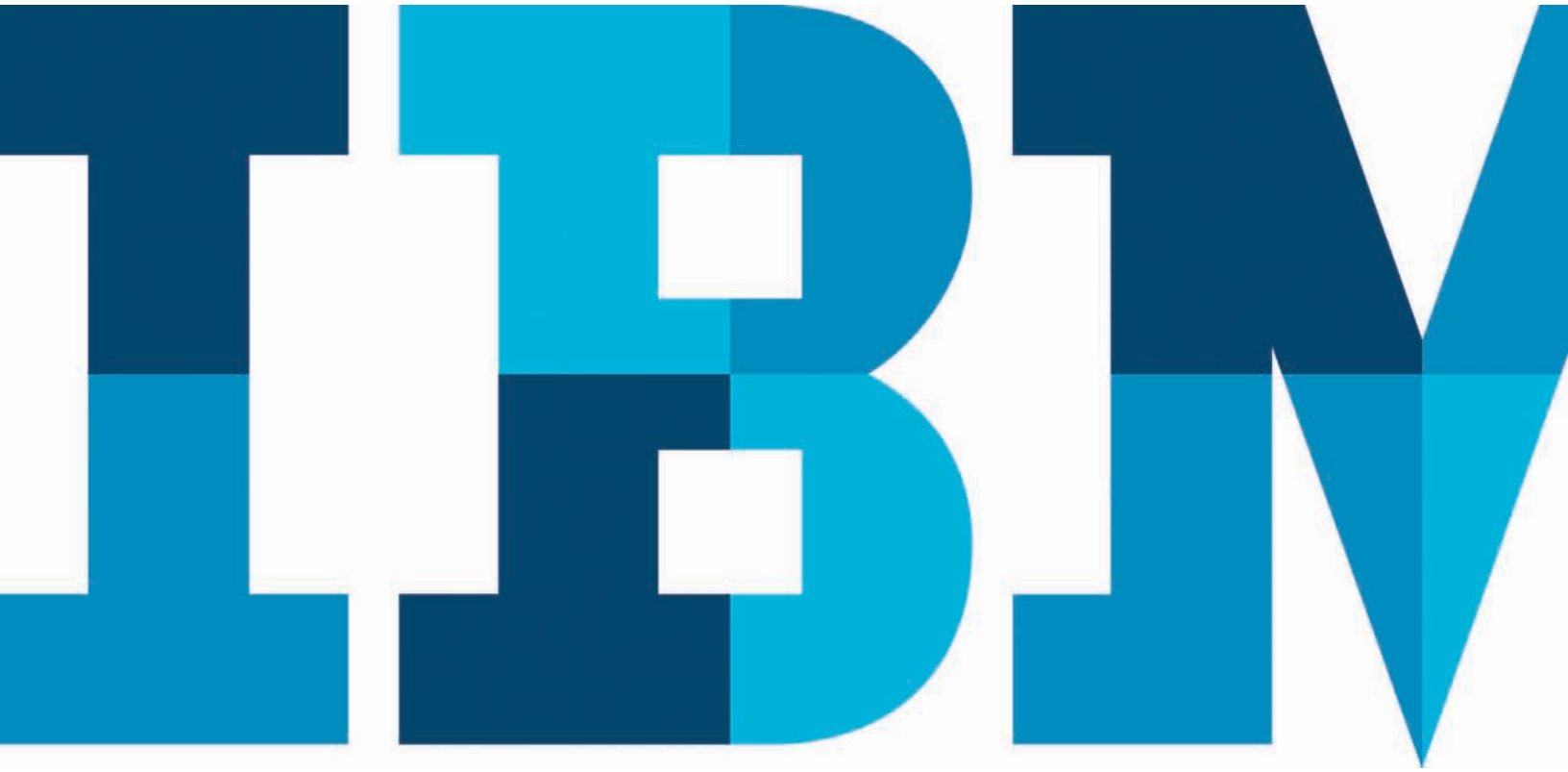
IBM Software
Thought Leadership White Paper

January 2013

Extending security intelligence with big data solutions

Leverage big data technologies to uncover actionable insights into modern, advanced data threats



IBM

Introduction

Sophisticated cybercrimes and advanced persistent threats are occurring at an alarming rate. Aided by new attack techniques, increased financial support and the ease of exploiting social connections, attackers are having more success than ever before. Traditional security solutions are no longer sufficient to defend against these escalating threats.

IBM® Security QRadar® uses big data capabilities to help keep pace with advanced threats and prevent attacks before they happen. It helps uncover hidden relationships within massive amounts of security data, using proven analytics to reduce billions of security events to a manageable set of prioritized incidents.

Forward-leaning organizations are exploring custom analytics that use additional big data technologies on a variety of unstructured data sources including email, social media feeds, business transactions and full network packet payloads. To meet this demand, IBM is integrating industry-leading security intelligence capabilities with the world-class analytics capabilities of IBM InfoSphere® BigInsights™ and related big data software and services. The combination offers a comprehensive solution—a security intelligence platform designed to detect and prioritize threats in real time, together with a mature Hadoop-based solution for custom data mining and analytics.

Understanding and identifying advanced threats

Advanced threats have become one of the IT security industry's most discussed topics. Today, organized teams pursue specific targets via well-orchestrated, patient, long-running attacks, often using highly customized malware and tactics. For example, the attacker may infiltrate a trusted partner and then load malware onto the target's network. The malware may be tailored to infect only the target organization to prevent identification by security vendors.

Attackers also perform extensive reconnaissance of spear-phishing targets and then compromise their accounts via social engineering tactics. Adversaries often exploit zero-day vulnerabilities to gain access to data, applications, systems and endpoints, and they communicate over a variety of channels to exfiltrate data from the targeted organization.

To combat these and other sophisticated threats, organizations must adopt new approaches that help spot anomalies and subtle indicators of attack. Doing so requires collecting and analyzing data from the security infrastructure and beyond—including traditional log and event data as well as network flow data, vulnerability and configuration information, identity context, threat intelligence and more. In short, security is becoming a big data problem.

Security intelligence: A proven big data approach to security

Security intelligence is the continuous real-time collection, normalization and analysis of data generated by users, applications and infrastructure. It integrates functions that have typically been segregated in first-generation security information and event management (SIEM) solutions, including log management, security event correlation and network activity monitoring.

Data collection and analysis goes well beyond traditional SIEM, with support for not only logs and events, but also network flows, user identities and activity, asset profiles and configurations, system and application vulnerabilities, and external threat intelligence within the single warehouse, as illustrated in Figure 1.

Expanding visibility with IBM QRadar Security Intelligence Platform

The IBM QRadar Security Intelligence Platform is a big data platform designed from the ground up to deliver the benefits of next-generation SIEM technology. It is designed to expand

visibility into network, virtual, user and application activity to help provide actionable intelligence into potential security offenses across the organization.

Major enterprises use QRadar solutions to collect and correlate billions of events and network flows per day in deployments that span multiple locations and connect previously siloed operational groups. Examples include:

- A Fortune 100 telecommunications provider collects and monitors one million events per second—more than 85 billion events per day—to aid security and regulatory compliance across its global operations.
- A global energy company monitors six million card swipes and correlates two billion events per day to aid compliance with standards of the North American Electric Reliability Corporation (NERC) and the Payment Card Industry Data Security Standard (PCI DSS). Real-time analysis provided by QRadar solutions determines the 25 to 50 priority incidents that matter each day—for a roughly 40-million-to-one data reduction ratio.

Several elements make QRadar solutions an ideal approach to help combat advanced threats:

- **Interoperability and scalability:** QRadar solutions support more than 400 information sources and help provide a unified architecture for collecting, storing, analyzing and querying log, threat, vulnerability and risk-related data. A purpose-built database helps provide scalability and performance tuning, allowing organizations to search millions of events with sub-second response.
- **Pre- and post-exploit insights:** QRadar solutions gather and prioritize information about existing security gaps to help prevent breaches; they help identify suspicious behavior already taking place within the network to help detect breaches.
- **Anomaly detection capabilities:** QRadar solutions create a baseline of current activity to identify deviations from normal behavior. They then determine which deviations are meaningful in order to help detect attacks in progress.

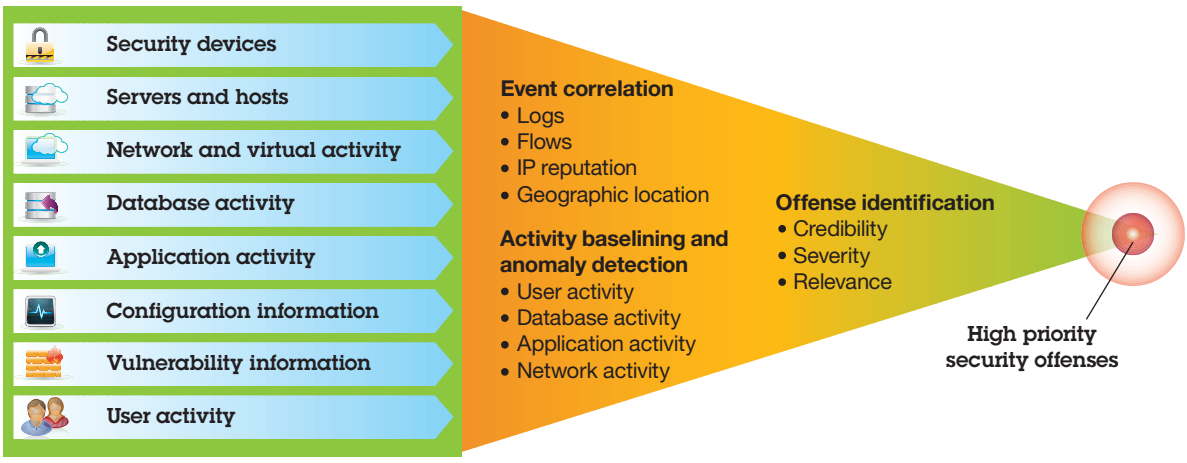


Figure 1. The scalable IBM QRadar Security Intelligence Platform captures data across a broad range of feeds, reducing it to a manageable list of offenses using pre-existing and customer-defined rules.

- **Real-time correlation and analysis:** QRadar solutions correlate massive data sets in real time, helping allow for earlier and more accurate detection of advanced attacks.
- **Reduced false positives:** QRadar solutions can quickly help detect compromises and de-prioritize unusual yet benign activity, reducing the time analysts spend investigating potential breaches.
- **Forensic capabilities:** QRadar solutions provide a single-console view of log data, network traffic and other security telemetry across thousands of systems and resources—helping to ease the burden on the security and network staff who have to rapidly assess the source and impact of a breach.
- **Flexibility:** Effective approaches to defending against advanced attackers should support frequent change in the IT environment and threat landscape. QRadar solutions help make it easier to add data sources, create and tune analytics, create new user views and reports, and expand and evolve the overall deployment architecture.

Leveraging parallel processing, high-speed ad hoc querying capabilities and analytics to correlate and refine millions of security events into a manageable set of priority offenses, the QRadar platform is a leader in the use of big data for IT security.

Using additional big data tools to solve new problems

Forward-leaning organizations are turning to big data platforms such as those based on Hadoop to help solve advanced security challenges. The type of analysis these platforms provide typically uses historical baselines, statistics and visualizations to uncover evidence of past fraud or security breaches. Examples include:

- Communications providers are correlating millions of global DNS requests, HTTP transactions and full packet information to identify malicious communications associated with botnets.
- International financial services companies are gaining new approaches for uncovering fraud by correlating real-time and historical account activity, and by using baselines to spot abnormal user behavior, unlikely application paths and suspicious transactions.
- Organizations are using linguistic and predictive analytics to profile email and social networking communications and to identify suspicious activity, triggering proactive measures before an incident takes place.

Big data analytics employed in these use cases must store, process and analyze a wide variety and volume of structured, semi-structured and unstructured data that is not currently analyzed in today's security solutions, as illustrated in Figure 2.

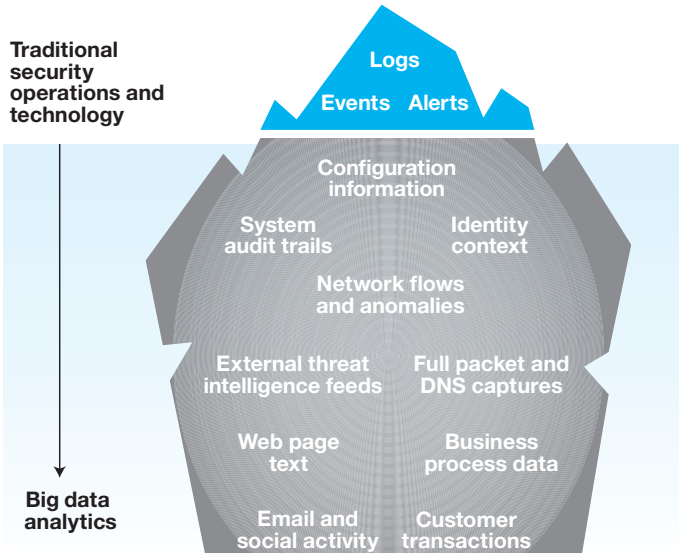


Figure 2. The variety and volume of data is driving new big data use cases to help enterprises maintain security.

Extending security intelligence with IBM InfoSphere BigInsights

IBM combines the expert security capabilities of the QRadar Security Intelligence Platform with advanced analytics technologies including IBM InfoSphere BigInsights, a Hadoop-based platform that helps organizations discover insights hidden in large volumes of data. QRadar solutions perform real-time correlation and reporting for rapid threat and risk response and then send enriched security information to InfoSphere BigInsights for additional analysis.

InfoSphere BigInsights can consume and analyze immense amounts of data from unstructured and semi-structured sources, accommodating both the variety and volume of data needed for advanced security use cases. InfoSphere BigInsights can help improve the accuracy of analysis over time and feed insights back to QRadar, 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, as illustrated in Figure 3.

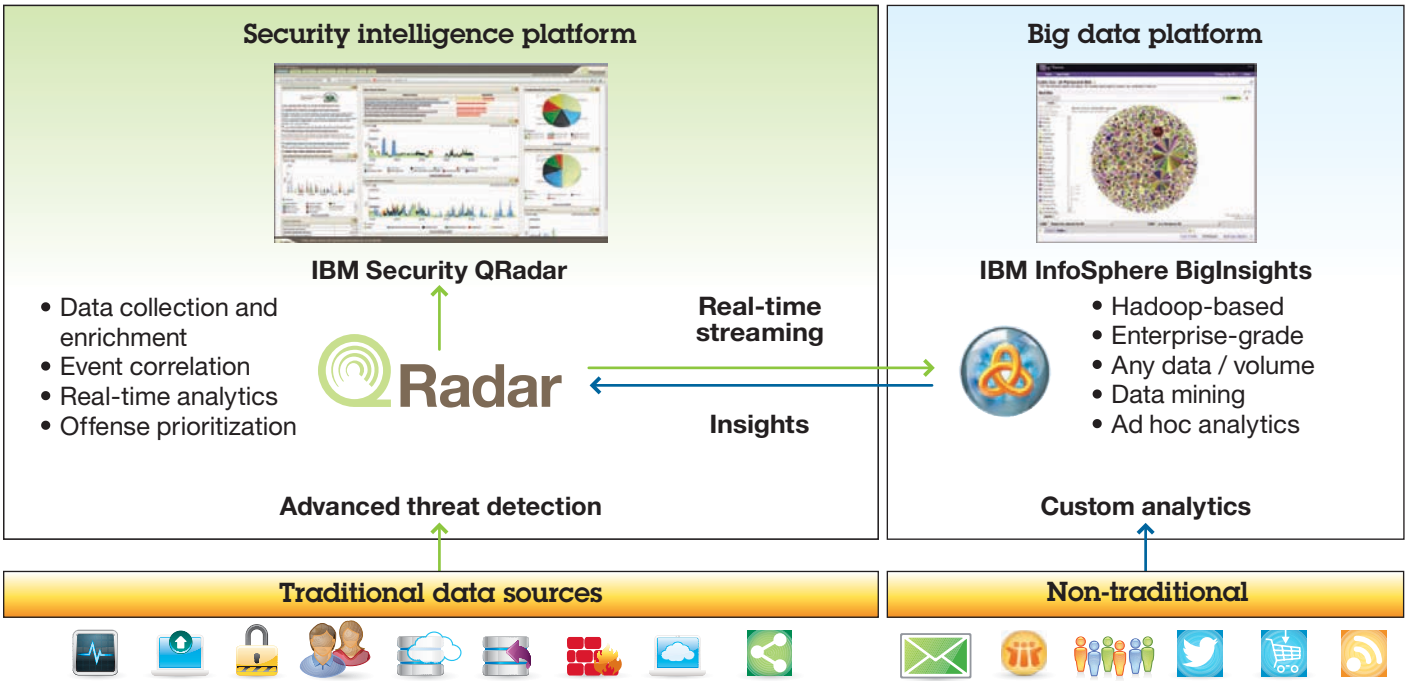


Figure 3. InfoSphere BigInsights uses adaptive analytics to help extend the QRadar Security Intelligence Platform with big data capabilities.

InfoSphere BigInsights complements QRadar solutions with big data processing features including:

- **Sophisticated text analytics** with a vast library of extractors enabling actionable insights from large amounts of native textual data
- **A machine data analytics accelerator** to ingest, parse and extract a wide variety of machine data with faceted search for easy navigation, discovery and visualization
- **Adaptive MapReduce,** adapting to user needs and system workloads automatically to help improve performance and simplify job tuning without the need for users to understand and manipulate the many tuning knobs in Hadoop
- **IBM BigSheets,** a spreadsheet-like tool that allows users to explore InfoSphere BigInsights collections and discover new insights without writing code

InfoSphere BigInsights includes advanced analytics and user interfaces for the non-developer security analyst. The solution does not require schema definitions or data preprocessing, and allows for structure and associations to be added on the fly across information types. The platform runs on commonly available, low-cost hardware in parallel, supporting linear scalability on commodity hardware. Depending on the use case requirements, InfoSphere BigInsights can be augmented by additional IBM big data technologies.

Big value from big data: Internet-scale botnet discovery

Organizations can significantly improve their security posture by identifying botnet-infected hosts and limiting communications to command-and-control hosts.

Technical challenges

Botnets can cause infected hosts within a network to leak very small amounts of data by requesting new commands and sending information sporadically over a period of time. Detecting this type of random, infrequent activity is difficult and involves monitoring large volumes of high-velocity data with rapidly changing identifiers such as DNS traffic and other protocols used for command and control.

Additional IBM big data technologies

- **IBM InfoSphere Streams:** Parallel processing techniques for performing complex analytics on massive volumes of data in motion, such as text, images, audio, voice, voice over IP (VoIP), video, web traffic and email content at rates up to petabytes per day. This solution is available for highly custom, high velocity use cases that require analysis in milliseconds rather than seconds.
- **IBM SPSS® Modeler:** A data-mining workbench that helps the data analyst build predictive models quickly and intuitively using structured and unstructured data.
- **IBM i2® Intelligence Analysis Platform:** A powerful visualization tool that helps analysts discover trends and disseminate actionable threat information to the enterprise.
- **IBM PureData™ System:** A high-performance appliance that helps simplify and optimize performance of data services for analytic applications, helping enable very complex algorithms to run in minutes, not hours.

The IBM solution

- Use QRadar solutions for native collection of network flows help to identify botnets, detect anomalies in real time and correlate malicious activity against IBM X-Force® global threat intelligence.
- Collect virtually all DNS transactions across the enterprise using InfoSphere BigInsights and apply custom analytics to help identify suspicious domain names used by botnets. Analyze years of historical data to help detect infected hosts and past intrusions.
- Integrate findings from InfoSphere BigInsights—including command-and-control domains and assets requiring remediation—into QRadar to build real-time correlation rules to help in spotting new intrusions.

Big value from big data: Full-spectrum fraud detection

Organizations lose substantial revenue to fraudulent claims, account takeovers and invalid transactions each year. Despite the magnitude of the problem, many organizations are unaware that fraud is being committed against them.

Technical challenges

Fraud analysis involves looking for anomalies and behavior patterns and building a profile of normal activity. Security teams and fraud investigators need deep access to this information and the ability to parse unstructured text to understand discrepancies in customer transactions, claims and other behavior.

The IBM solution

- Employ QRadar solutions to collect, normalize and enrich application and user access logs and transaction data, searching for anomalies in real time and sending post-processed information to InfoSphere BigInsights.
- Use InfoSphere BigInsights to perform custom analytics on transactions and to baseline petabytes of account activity over months and years—sending insights back to the QRadar platform to detect fraudulent activity as it occurs.
- Extend this information to the IBM i2 Intelligence Analysis Platform for link analysis, visualization and dissemination to help fraud analysts conduct investigations and communicate findings with others.

Big value from big data: Comprehensive insider threat analysis

Insider threats and data loss are major risks for any organization—and the stakes are high, especially as repositories of customer/personal information expand.

Technical challenges

Most security technologies look only for specific patterns or short term profiles of "normal" activity for a user or application. To accurately detect insider threats, organizations may need to analyze months or even years of network traffic, IP addresses and URL destinations—as well as broader inter- and intra-company communications—to better understand people-to-people linkages and what constitutes risky behavior.

The IBM solution

- Use QRadar solutions to break down data silos and correlate real-time system and user activity—helping to uncover risky behavior by ordinary and privileged users accessing sensitive information.
- Leverage InfoSphere BigInsights for enhanced analytics useful for discovering and investigating high-risk behavior with their ability to look for patterns indicative of abnormal employee activity.
- Share findings with existing identity and access management systems—such as IBM Security Privileged Identity Manager—to help take corrective action on suspicious users and control shared access to sensitive information.

Conclusion

A big data security analytics solution must ingest data provided by a large variety of security data feeds from within the enterprise, as well as unstructured and semi-structured data from inside and outside the enterprise. It must also adapt to the changing threat landscape, provide a holistic view of the enterprise environment and drive actionable intelligence to protect against both known and unknown threats.

Using IBM Security Intelligence with Big Data, security organizations can analyze more data more flexibly, and gain more accurate results. By analyzing structured, enriched security data alongside unstructured data from across the enterprise, the IBM solution helps find malicious activity hidden deep in the masses of an organization's data, for advanced threat and risk detection.

For more information

To learn more about IBM security solutions for big data, contact your IBM representative or IBM Business Partner, or visit:
ibm.com/security

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
January 2013

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

All client examples cited or described are presented as illustrations of the manner in which some clients have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual client configurations and conditions.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED "AS IS" WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided. Hadoop is not an IBM product or offering. Hadoop is sold or licensed, as the case may be, to users under Apache's terms and conditions, which are provided with the product or offering. Availability, and any and all warranties, services and support for Hadoop is the direct responsibility of, and is provided directly to users by, Apache.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

Statement of Good Security Practices: IT system security involves protecting systems and information through prevention, detection and response to improper access from within and outside your enterprise. Improper access can result in information being altered, destroyed or misappropriated or can result in damage to or misuse of your systems, including to attack others. No IT system or product should be considered completely secure and no single product or security measure can be completely effective in preventing improper access. IBM systems and products are designed to be part of a comprehensive security approach, which will necessarily involve additional operational procedures, and may require other systems, products or services to be most effective. IBM does not warrant that systems and products are immune from the malicious or illegal conduct of any party.

Please Recycle

WGW03020-USEN-01