CHAOSSEARCH

A SCALABLE SECURITY DATA PLATFORM

The Foundation of a Modern Enterprise's Security Operation

While log data has always played an important role in the cybersecurity framework, the ever-increasing frequency, cost and sophistication of cyber-attacks have made log management and analytics a mission critical responsibility of the security operations (SecOps) team.

TechTarget Custom Media

CONTENTS

INTRODUCTION

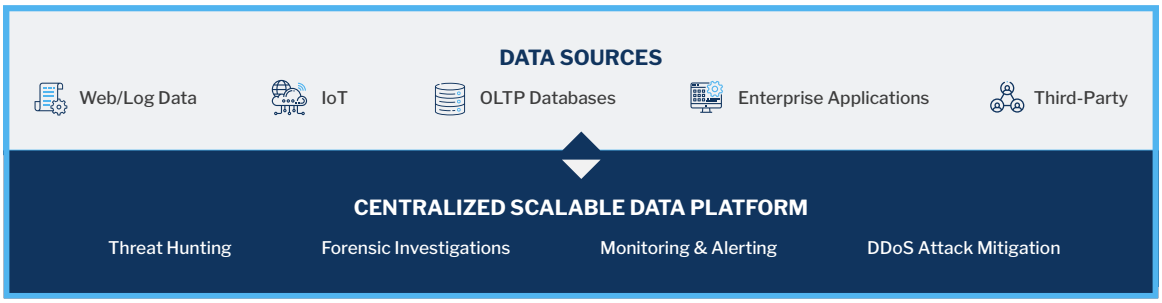
While log data has always played an important role in the cybersecurity framework, the ever-increasing frequency, cost and sophistication of cyber-attacks have made log management and analytics a mission critical responsibility of the security operations (SecOps) team.

Indeed, most organizations see log analytics as a foundational element of their overall cyberdefense strategy.

However, the heightened importance of data analytics, coupled with the explosion of data growth of recent years, are necessitating a dramatic shift in how enterprise SecOps teams collect, manage and use log data to power their operations and improve the overall security posture of their organizations. Traditional log management systems cannot scale efficiently to meet the needs of today's SecOps teams which rely on access to massive data sets, from a wide range of sources, including long-term historical data. These limitations are driving the need for a massively scalable data platform that overcomes inherent limitations of the traditional approach, allowing SecOps teams to secure the foundation of their overall cybersecurity architecture.

THE VALUE OF LOG DATA ANALYTICS TO SECURITY OPERATIONS

Log data and the analytics it enables provide details on extreme traffic, unauthorized access, suspicious changes and many other pointers used to identify potential threats. These indicators are at the heart of many core SecOps activities, including detection and alerting, forensic investigations, threat hunting, insider threat detection, and DDoS attack prevention. When all relevant data is stored in a centralized repository, it can be the "single source of truth" about both what has happened and what exists right now across an organization's IT environment. Thus, centralization of data from a wide range of sources is crucial as it allows analysts to conduct complex queries and correlations from a variety of data streams.



THE RISE OF BIG DATA NECESSITATES IMPROVED LOG DATA AND ANALYTICS SOLUTIONS

The amount of log data generated every day has been increasing rapidly for some time, outpacing most organizations' ability to capture, organize and utilize all of it. This has resulted in some troubling realities for IT and SecOps teams and may lead to significant vulnerabilities.

IT and SecOps professionals recognize that log data that goes uncaptured cannot be used to assess threats or track suspicious activities. Nor can it aid in uncovering potential threats. The inherent vulnerabilities in this status quo are unacceptable.

The need for managing ever-growing volumes of log data continues to increase, with no end in sight. As organizations add more and different security tools, these new defenses will generate their own unique log data, which must be included with the existing data sets. In some cases, new cybersecurity solutions will add larger data feeds than were common in the past.

In addition, log data from other systems, such as content delivery networks, will become part of the overall data lake. Given that a single new corporate application can generate millions of log files per day, it becomes clear that as the overall IT environment grows larger and more complex with new services, apps and devices, there is a multiplying effect on the amount of new log data generated each day.

For example, a recent customer survey found that on average:

60%	80%	90%
⅓ of organizations capture less than 60% of their log volume	IT teams capture slightly less than 80% of their log volume	Half of survey respondents said their ideal coverage rate is 90% or more

EFFECTIVE SECURITY DEMANDS A NEW SCALABLE AND COMPREHENSIVE DATA ANALYTICS SOLUTION

The ability to collect, analyze and store very large sets of log data is clearly a requirement moving forward. Organizations that don't meet this challenge will be more vulnerable, and any security event or threat will be harder to identify and remediate. That is an unacceptable state. Legacy log management platforms, including the widely used ELK stack—the open source solution that combines Elasticsearch, Logstash and Kibana—were not designed to meet the scaling requirements of the modern IT organization and have become untenable, driving interest in alternative solutions.

Scaling Up Data Ingestion

As noted, the amount of daily log data being generated is growing by orders of magnitude. The inclusion of more sources of log data and the need to keep it "live" longer make it necessary to support 25 terabytes or more of daily log data. As noted, the volume of log data that companies are generating on a daily basis is growing by orders of magnitude, and this growth is accelerating as more sources of business-critical data are identified. It's crucial, therefore, that a company's log management system can easily and efficiently scale to accommodate the increasing data loads. If it can't, business insights and operations will surely suffer.

> When legacy platforms are limited to capturing 65% of daily log data, with that number falling, it creates a crisis.

Ingesting Data in its Native Form

To choose the right platform to scale up log data storage effectively requires some forward thinking. One of the most important factors is ensuring that adding new data sources is simple and seamless. As indicated, important log data is going to come from many new sources. The process for adding a new data source can't take days of a security engineer's time. Along with simplifying the addition of new sources, the new log data management tool must have sufficient headroom for growing log data and possess the agility to incorporate diverse new sources.

Such agility can't be overlooked. The most important manifestation of agility is the ability to ingest data in its native form without any parsing or transformation necessary for inclusion. The ability to ingest data in native form improves speed, eliminates the potential for mistakes or issues related to parsing/transformation, and eliminates the risk that some data is lost during the ingest process.

The optimal approach is to store the log data in a security data lake that leverages cloud object storage as the target for all log data from all sources. A platform that connects to and indexes the full repository can then enable existing search and analytics tools to access the data. This approach solves the problem of having the analytics platform act as the data custodian, responsible for organizing and managing all of the data. Serving as the data custodian of a massive, rapidliy growing repository gives rise to scalability, performance and reliability problems.

INCREASING THE DATA RETENTION PERIOD IS AN IMPORTANT COMPONENT OF SCALING UP

The focus on scaling up capacity for log data is essential, but fully solving the problem demands more. The second action necessary is increasing the amount of time that log data is retained. ChaosSearch has found that many IT and SecOps teams retain only two to four weeks of log data. This is not nearly enough; a more appropriate period of live data retention should be closer to six months. The most common reason for this short retention period is the inability of their existing log data management system to serve as the very large data store that results from long term retention periods.

Retention Periods Need to be Much Longer for Numerous Reasons, Including:

Improved ability to find advanced persistent threats (APTs): The persistent nature of this type of attack makes it necessary to look back further in time to identify the pattern of attack commonly used by APTs. The probing and infiltration stage can take months, and that mandates the need to review data collected over a long period of time.

Support for threat hunting: As threats become more clever and harder to spot, SecOps teams benefit from having more data to analyze and drive pattern recognition. As threats attack different aspects of infrastructure at different times, having both more data and data that reflects a longer time frame improves the analytics or machine learning/AI that supports threat hunting.

Alignment with dwell time metrics: The median dwell time for an incursion is 56 days, based on the latest research from FireEye.[1] Unfortunately, if the log data covers only a 14-day period, any incursion that happened weeks before won't be easily identified. Increasing the retention period to go beyond average dwell time improves the odds of finding incursions that occurred weeks or months ago.

[1] "FireEye Mandiant M-Trends 2020 Report Reveals Cyber Criminals Are Increasingly Turning to Ransomware as a Secondary Source of Income," FireEye, Feb. 20, 2020

ADDRESSING THE LIMITATIONS OF LEGACY LOG MANAGEMENT SOLUTIONS

A scalable log management system that enables high volumes of daily ingest eliminates the need for complex data transformations and delivers the long-term data retention required to replace the commonly deployed ELK stack.

Put simply, it's time to move past legacy approaches and deploy a scalable log data solution that enables the elimination of ELK cluster sprawl. The space limitations of ELK data stores have two suboptimal outcomes: cluster sprawl and so-called sharding of ELK data stores. Both add complexity and operational issues. Worse, it becomes harder to gain a comprehensive picture of events, thereby reducing an organization's overall security posture. With a better, more scalable solution, both problems can be eliminated.

SUMMARY AND KEY TAKEAWAYS

Analysis of log data is a foundational element of effective cybersecurity, and its importance is on the rise as the frequency and sophistication of cyber-attacks increase. And as both IT infrastructure and the number of security tools increase, the amount of log data is growing exponentially. However, many legacy solutions for managing log data fall short when they are required to scale effectively, necessitating a new approach. A massively scalable data platform that scales up in terms of its ability to ingest data and scales out to enable long term data retention is required.

ChaosSearch delivers this type of next-generation scalable data platform. Because of its unique approach, which indexes data in cloud object storage without assuming custody of the data, the ChaosSearch Data Lake Platform can scale up to enable any required daily ingest rate (currently, their largest customer processes north of 30 TBs per day), and can manage the necessary capacity to allow for very long data retention periods (multi-year if needed). Importantly, ChaosSearch's solution also makes it very easy to add new data sources without requiring any parsing or schema changes, making them quickly available for use by the SecOps analysts.

For more information on deploying a more modern and capable platform for log data management and analytics, please go to: www.chaossearch.io

ABOUT CHAOSSEARCH

ChaosSearch empowers data-driven businesses like Blackboard, Equifax, and Klarna to Know Better™, delivering data insights at scale while fulfilling the true promise of the data lake. The ChaosSearch Data Lake Platform indexes a customer's cloud data, rendering it fully searchable and enabling data analytics at scale with massive reductions of time, cost and complexity. The Boston-based company raised $40M Series B in December 2020 and is hiring to support its hyper growth.

For more information, visit ChaosSearch.io or follow us on Twitter @ChaosSearch and LinkedIn.

info@chaossearch.com | www.chaossearch.io