SentinelOne

Transforming Big Security Data into Strategic Advantage with Singularity AI SIEM

Cloud-native data platform turns big security data into a strategic advantage for defenders

White Paper

Table of Contents

Introduction

Cyber defenses generate an unprecedented, sometimes overwhelming, wealth of data and visibility for defenders. Every day, enterprises generate terabytes of data from endpoints, network devices, applications, and cloud services. This massive explosion in data, if harnessed correctly, provides the key to defending against increasingly sophisticated cyber threats, driving successful outcomes such as:

Proactively protecting the business with the latest threat, vulnerability, and security posture data.

Detecting sophisticated attacks across past and present threats within an environment.

Reducing mean-time-to-detect, and mean-time-to-respond to incidents, minimizing impact.

Recover faster by arming responders with the deep context needed to restore systems and get back to business.

The future of cybersecurity rests on the power of advanced data platforms, and their ability to unlock new insights and workflows for defenders. Harnessing this data, however, is no small task. Legacy Security Information and Event Management (SIEM) systems have struggled to keep pace with the evolving cybersecurity landscape. These traditional platforms are often overwhelmed by the sheer volume and complexity of today's security data. According to one Gartner report, an alarming 60% of SIEM and Data Lake projects fail, underscoring the need for a new approach in managing security data.

In this report, we'll explore the limitations of legacy data solutions and the transformative potential of Singularity AI SIEM. With AI SIEM, security teams are turning their security data into a strategic asset that drives better outcomes from Security Operations Center (SOC) workflows and improves defenses against the cyber threats of tomorrow.

Legacy Security Data Solutions Don't Hold Up in the Post-EDR World

Perhaps no development in cybersecurity has had a bigger impact on data architecture than Endpoint Detection and Response (EDR). EDR provides a wealth of detailed information, capturing everything from process activities and network communications to registry changes and file interactions. This rich data captures a high-resolution picture of events and security incidents from the endpoint perspective, offering deep insights for defenders.

This data has proven invaluable for security teams, but it rarely stands on its own in the modern SOC. While EDR solutions excel when it comes to monitoring and protecting traditional internal and perimeter endpoints, the modern digital landscape extends far beyond these boundaries. Organizations operate in a distributed ecosystem that includes cloud services, SaaS applications, and remote infrastructure, where EDR's visibility quickly drops off.

In the post-EDR world, the need for comprehensive visibility and control remains critical, regardless of where an organization's assets, processes, or data live. Security teams need a centralized platform that aggregates security data across the full spectrum of security solutions if they are going to overcome the silos that create friction and increase risk.

While SIEM systems might seem like the natural choice for integrating these diverse data sets, they fall short in practical application. On average, a single EDR tool can generate upwards of 2TB of data per day for every 1,000 endpoints. Ingesting EDR data at this scale into a SIEM is impractical for a number of reasons:

- **SIEMs are Prohibitively Expensive**
  Most SIEM solutions are priced based on the amount of data ingested; bringing full EDR telemetry into a SIEM can increase costs dramatically. Organizations often compromise security effectiveness when they are forced to discard valuable data due to financial constraints.

- **SIEMs are not Designed for EDR-scale Data**
  Legacy solutions like SIEMs are built on data architectures that date back decades. Originally designed for the data centers of the past, they have been lifted and shifted into the cloud with little fundamental innovation. They incorporate antiquated flat relational data schemas and have indexing requirements that bring significant overhead in terms of resource utilization and management expenses, leading to bottlenecks and inefficiencies.

- **Tiered Storage Creates Friction**
  Expensive legacy data architectures drive the need for complex tradeoffs when it comes to data storage. In order to reduce storage costs, historical data may be relegated to cold or frozen storage tiers, which may require cumbersome processes to reingest or rehydrate when it's needed, introducing friction and delays to incident investigations. As a result, query response times can fail to keep up with the fast-paced needs of security operations teams.

- **SIEMs Erase Context**
  When EDR data is shoehorned in to fit a SIEM data model, it must be flattened and normalized to fit into the native schema of the underlying platform. In the process of flattening EDR data, connections between individual data elements and the rich narratives critical for understanding sophisticated attacks are often lost, erasing much of the inherent value.

- **SIEMs Require Highly Skilled Security Engineers**
  The complexity and inefficiency of legacy solutions necessitate highly skilled experts to parse and normalize data, and to engineer queries used by analysts to uncover and understand emerging threats.

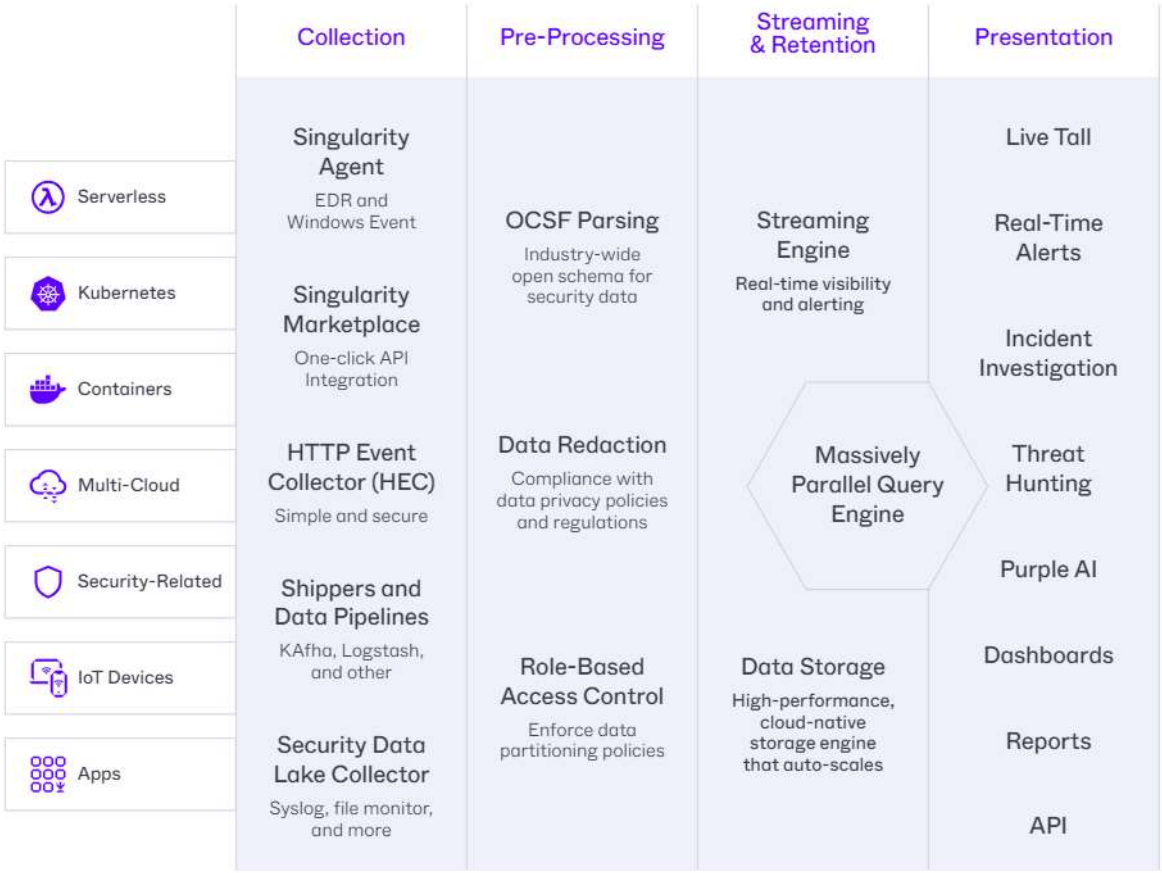
Legacy data solutions are relics of a bygone era in cybersecurity, not suited to the demands of modern security operations. The challenges they present—in terms of architecture, operational efficiency, resource requirements, and cost—underscore the need for a new approach. This approach must be capable of seamlessly integrating with and exploiting the rich data generated by advanced EDR systems as well as the broader security ecosystem.

Introducing Singularity AI SIEM

Singularity AI SIEM is a cloud-native, modern solution that empowers security teams by unleashing the power of their security data. It allows organizations to merge the full spectrum of security data sources into a cohesive and comprehensive security data lake, offering a scalable, efficient, and cost-effective alternative to legacy SIEM platforms.

Figure 1 provides a high-level overview of the AI SIEM architecture. Let's review the components to see how they come together to deliver improved security outcomes.

Figure 1: SentinelOne Security Data Lake Architecture



Collection

The first step in operationalizing security data is ingesting it into the platform. AI SIEM provides a variety of collection methods to meet the needs of a broad range of data sources.

- **Singularity Agent**
  EDR data remains the foundation of many security workflows. The SentinelOne Singularity Agent seamlessly collects Windows Event Logs and EDR data and streams these data to the AI SIEM in near real time.
- **Singularity Marketplace**
  SentinelOne Singularity Marketplace is an ecosystem to help customers extend the Singularity Platform with pre-built, validated integrations. With Singularity Marketplace, organizations can ingest data and orchestrate responses across dozens of security solutions into AI SIEM with a single click.

- **HTTP Event Collector (HEC)**
  HEC allows applications to securely send data to the AI SIEM via the HTTP protocol. HTTP collection is fast, efficient, and easy to integrate into custom applications.

- **Shippers and Data Pipelines**
  Many organizations and applications use third-party data pipelines such as Kafka, LogStash, and Cribl to move data across a distributed application infrastructure. AI SIEM can tap into these pipelines to easily capture data that's relevant for security use cases.

- **AI SIEM Collector**
  There are a number of common default methods that applications use to store and share events and logs, including syslog and log files among others. The AI SIEM Collector provides the flexibility to collect and operationalize this data.

Streamlined Migration

Many legacy SIEMs, especially Splunk, leverage HTTP Event Collection to ingest data from a wide range of applications. AI SIEM's HEC makes it easy for many organizations to migrate high-volume data sources away from expensive legacy SIEMs, reducing costs and improving access in a matter of minutes.

Parsing

Data parsing enables interpretation and normalization of diverse data formats from varied sources. Effective parsing ensures that AI SIEM can correlate events accurately, facilitating timely threat detection and response. At the core of AI SIEM's data architecture is the Open Cybersecurity Schema Framework (OCSF).

OCSF is an industry-wide open schema that facilitates the breakdown of data silos, speeds up integration, and simplifies interoperability across different security tools. By adopting OCSF, AI SIEM ensures efficient and future-proof ingestion and rule creation, allowing for a write-once, apply-everywhere approach. This standardization also eliminates vendor lock-in, guaranteeing data portability and flexibility.

Example: imagine you are a threat hunter with new intelligence about an adversary who is using a specific set of IP addresses as launch points for an emerging attack. You'd like to know if anything in your organization has interacted with this hostile infrastructure, and to create a rule to alert you if it happens in the future.

You have logs from a variety of sources, perhaps including Okta for identity, Fortinet firewalls, and SentinelOne Singularity Platform, among many others. OCSF parsing allows you to create a single query that runs in seconds to identify malicious activity across all your data in seconds.

Once your historical search is complete, the query can be easily converted to a rule that alerts the security team and takes action in real time to mitigate risk of a damaging breach.

Go Deeper with OCSF

To learn more about how AI SIEM uses OCSF to improve security efficiency and effectiveness for real-world organizations, see our blog, Simplifying the Security Analyst Experience with Open Cybersecurity Schema Framework.

Streaming Engine

Once data is parsed and normalized, it's time to put it to use. Processed data is fed into the AI SIEM platform's streaming engine, which allows for real-time visibility and alerting, ensuring timely responses to emerging threats. The streaming engine applies security rules and AI models to identify malicious and suspicious activity as it happens, and to generate alerts to drive fast response. It also makes processed data immediately available for review via Live Tail, so administrators and analysts are armed with the information they need to track threats and troubleshoot issues in real time.

Storage and Queries

In parallel with the AI SIEM streaming engine, processed data is also stored for future use in detecting and investigating threats, threat hunting, compliance, and other use cases. AI SIEM's storage architecture represents a significant departure from traditional systems, eliminating the need for indexing. Parsed logs are stored in columnar format in a modern, cloud-native data store, reducing the complexity and overhead associated with data management. The AI SIEM storage is paired with a massively parallel query engine, leveraging container technology that breaks down complex queries into manageable tasks, aggregating results swiftly.

This architecture enables AI SIEM to run many complicated computations simultaneously, significantly outpacing legacy SIEM and XDR systems. Analysis shows that 96% of AI SIEM queries complete and return answers to analysts in under 1 second, as much as 10x faster than legacy SIEM and XDR providers. This level of performance not only speeds up interactive queries that an analyst might make during an investigation, but also powers real-time dashboards, threat hunting, and unlocks the power of Purple AI.

Role-Based Access Control (RBAC) and Data Redaction

AI SIEM provides comprehensive data governance and security mechanisms, including effective RBAC and data redaction features. These tools enable organizations to partition data and responsibilities efficiently, ensuring that sensitive information is accessed only by authorized personnel. The platform supports multiple levels of tenancy, from global to site-level, facilitating granular control over data access and management.

Does Singularity AI SIEM Replace My SIEM?

Depending on your organization's needs, AI SIEM may offer a complementary and more efficient alternative to your current SIEM, and help the transition to XDR. Organizations who move workflows from legacy SIEM to AI SIEM enjoy a number of benefits:

- **Enhanced Data Ingestion**
  AI SIEM excels in handling high-volume data sets that are often cost-prohibitive for SIEM systems to process.
- **Extended Data Retention**
  Traditional SIEMs may impose restrictive data retention policies, often limited to 14-30 days for the highest-resolution telemetry due to cost and storage constraints. In contrast, AI SIEM supports longer retention periods in a cost effective manner, enabling organizations to maintain access to historical data for extended threat hunting and compliance purposes.

- **High-Performance Queries**
  AI SIEM's architecture allows for high-performance queries across large data sets, without friction caused by cold storage and data rehydration. This is a significant advantage over many SIEM systems, which can become bogged down when querying extensive historical data.

- **True Cloud Auto-Scaling**
  As data volumes grow, AI SIEM can scale dynamically, accommodating petabytes of data without requiring manual intervention from the operations team.

Still, many organizations find that it makes sense to maintain their legacy SIEM for use cases where the costs of migrating to a new platform are high, and the performance requirements are relatively modest. In these environments, SIEM and XDR can exist side-by-side, providing complementary capabilities.

Legacy SIEM	AI SIEM
Compliance Management: Collection and reporting of data to meet regulatory requirements such as GDPR, HIPAA, or PCI-DSS.	**Advanced Detection and Correlation:** Unifies EDR data with network, cloud, and identity, enabling sophisticated analytics needed to detect stealthy threats.
Security Dashboarding and Reporting: Delivering dashboards and reporting features to visualize and monitor security metrics and KPIs.	**Incident Investigations:** Graph data and Storyline quickly deliver the complete context, relationships and insights needed to understand and act in time to stop a damaging breach.
Log Management and Retention: Centralizing and managing logs from multiple sources for operational and compliance purposes.	**Threat Hunting:** Quickly search for indicators-of-compromise across all potential attack vectors with a single powerful query.
Forensic Investigation Support: Long-term historical data supports forensic investigations after a security breach or failure, helping teams understand the "how" and "why" behind an incident.	**Automated Response:** XDR platforms enable quick containment of threats by integrating with a variety of security controls to respond effectively in real time.

AI SIEM complements and extends beyond the capabilities of legacy SIEMs, significantly improving the security team's ability to detect and respond to threats quickly and effectively without disrupting existing workflows.

Putting Your Data to Work: Unlocking Security Outcomes with AI SIEM

AI SIEM offers a dynamic platform that transforms mountains of security data into new, actionable insights and workflows. By integrating AI SIEM into their security operations, organizations can unlock new investigative capabilities and threat response efficiency. Key use cases include:

- **Connecting Dots for Improved Investigations**
  AI SIEM's deep context, powered by graph analytics, empowers security teams to connect disparate events and data points across their entire digital footprint. AI SIEM's integrated approach means that an incident traced from an endpoint to Active Directory, then to a file server, and further to a cloud environment, can quickly be visualized and analyzed in a single view. This comprehensive visibility speeds up the investigative process, enabling analysts to understand the full scope of an attack and respond more effectively.

- **Natural Language Queries for Human Investigators and Threat Hunters**
  AI SIEM democratizes security data analysis by providing the performance needed to support natural language processing with Purple AI. Purple AI allows human investigators to interact with the Singularity Platform using plain English, drastically reducing the complexity and technical barriers traditionally associated with SOC workflows. Security analysts can simply ask questions, "show me all the logins from outside the US," and the system translates this into a structured query, executing it against the vast AI SIEM data repository in seconds. Purple AI then interprets the results, summarizing key findings and highlighting anomalies, thus streamlining the investigation process.

  Follow-up queries become more intuitive with AI-driven suggestions, such as asking for detailed actions of a suspicious user. AI SIEM, supercharged by PurpleAI, ensures that even non-technical staff can participate effectively in security operations.

- **Streamline Collaboration**
  AI SIEM provides a single, comprehensive source of truth for the entire security team, which is especially valuable during incident investigations. Signals and clues from natural language queries are automatically collected in investigation notebooks that can be easily shared among analysts as well as management and leadership. This makes it simple for multiple analysts to collaborate on an investigation, to hand off investigation responsibility across shifts, and to ensure that all stakeholders are kept informed without adding additional burden or reporting.

Go Deeper with Purple AI

To learn more about how Purple AI helps organizations to save time and resources by up-leveling every analyst with natural language queries and translation, see our blog, Transform SecOps with Purple AI.

- **Scaling to Power AI-Driven Security**
  The true potential of AI SIEM lies in its ability to scale and support advanced AI-driven security analysis. Complex queries like "What's the quickest exploit path into my environment?" demand extensive data about network resources, patch levels, configurations, and operational contexts. AI SIEM's robust infrastructure enables the rapid gathering and analysis of this information, providing AI with the necessary data to identify potential attack vectors and suggest proactive defenses.

  In a world where time is of the essence in mitigating cyber threats, the speed and efficiency of AI SIEM in processing and analyzing data can be the difference between a minor security incident and a catastrophic breach. The platform's ability to quickly parse through and make sense of vast amounts of data allows for real-time threat detection and response, thereby significantly reducing the risk window.

AI SIEM transforms the traditional role of data lakes from passive repositories to active, intelligent platforms that drive security operations. By enabling improved investigative capabilities, simplifying data interaction through natural language queries, and powering AI-driven security insights, AI SIEM provides the foundation for modern SOC workflows and ensures that organizations not only respond to current threats but also proactively prepare for future challenges.

Conclusion

The future of cybersecurity lies in the power of advanced data platforms. AI SIEM represents a fundamental shift in how organizations approach security data, offering a solution that is capable of handling the volume and complexity of modern cyber threats, turning this data into a strategic asset for defenders.

The integration of AI and machine learning technologies within AI SIEM has redefined the possibilities of threat detection and response. By harnessing these capabilities, organizations can predict and preempt security incidents with unprecedented accuracy and speed. The ability to conduct real-time, AI-driven analysis and to scale dynamically with the cloud ensures that AI SIEM not only meets the current demands of security teams but also evolves with the changing threat landscape.

As we look to the future, the role of data in cybersecurity will only grow in importance. The AI SIEM's approach to data management, with its emphasis on accessibility, integration, and actionable insights, is a blueprint for the next generation of security operations centers. It embodies the principles of efficiency, scalability, and intelligence that are essential for defending against the sophisticated cyber threats of tomorrow.

SentinelOne

Contact Us

techpartners@sentinelone.com
+1-855-868-3733

sentinelone.com

About SentinelOne

SentinelOne (NYSE:S) is pioneering autonomous cybersecurity to prevent, detect, and respond to cyber attacks faster and with higher accuracy than ever before. Our Singularity XDR platform protects and empowers leading global enterprises with real-time visibility into attack surfaces, cross-platform correlation, and AI-powered response. Achieve more capability with less complexity.

24_MKTG_Product_WhitePaper_009_Transforming_Big_Security_Data_with_AI_SIEM_r2_10282024