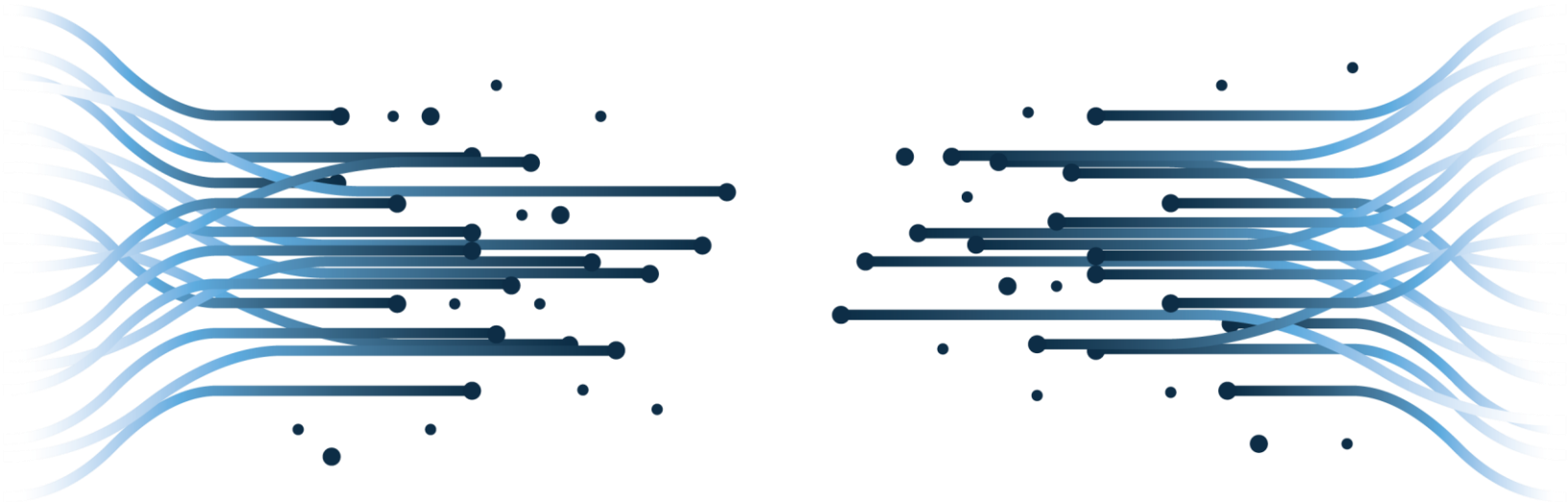
Optimizing SIEM Costs and Performance with Databricks Integration



NIS | Nationwide IT Services

www.nw-its.com

Nationwide IT Services, Inc. (NIS)
4025 Fair Ridge Drive Fairfax, VA 22033
www.nw-its.com 703-750-0453

Optimizing SIEM Costs and Performance with Databricks Integration

1. Introduction

As organizations scale their security operations, traditional Security Information and Event Management (SIEM) solutions like Splunk —while powerful—present growing challenges in cost, scalability, and advanced analytics. Nationwide IT Services (NIS) developed **NIS ARGUS™**, an integrated Databricks-based solution designed to address these pain points. **NIS ARGUS™** enables analytics and compliance while simultaneously reducing SIEM expenses, delivering enhanced real-time insights and alignment with federal cybersecurity standards.

2. Background

Federal agencies are under growing pressure to modernize cybersecurity capabilities while managing costs. Executive Order 14028 mandates the adoption of Zero Trust Architecture, multi-factor authentication, encryption, and enhanced event logging. In parallel, OMB Memorandum M-21-31 establishes a phased Event Logging (EL1 to EL3) model, requiring agencies to centralize, retain, and analyze logs with increasing maturity over two years.

To meet these requirements, agencies are increasingly turning to SIEM platforms such as Splunk, IBM QRadar, Microsoft Sentinel, and LogRhythm to manage their log ingestion, threat detection, and incident response workflows. Among these, Splunk stands out as a market leader, recognized by Gartner® for its powerful, feature-rich capabilities—including advanced analytics, machine learning (ML), and extensive integration support—which make it a preferred choice for large-scale enterprise deployments.

However, while Splunk delivers extensive functionality, many organizations are encountering significant challenges, particularly around licensing costs tied to data ingestion, performance at scale, and limited built-in support for custom Artificial Intelligence (AI)/ML analytics. These concerns have intensified following Cisco's acquisition of Splunk, which raises questions about long-term pricing, interoperability, and product direction.

To address these challenges, NIS developed **NIS ARGUS™**—a scalable, cloud-native integration between Databricks and Splunk designed to optimize log processing, reduce ingestion volumes, and unlock AI-driven analytics.

3. Industry Challenges with Traditional SIEM

Organizations across industries report the following recurring issues with Splunk:

- **High Data Ingestion Costs:** SIEM licensing based on data volume makes storing exponentially growing log data financially unsustainable.
- **Performance Bottlenecks**: Indexing and querying large data volumes introduce latency and slow threat detection.
- **Limited ML & Predictive Analytics**: Native AI/ML tooling in Splunk lacks maturity and flexibility.

- **Complex Configuration & Learning Curve**: Search Processing Language (SPL) and onboarding processes demand specialized expertise.
- **Inflexible Retention Policies**: Balancing storage costs with compliance needs is difficult.
- **Vendor Lock-in & Ecosystem Fragmentation**: Integrating with other tools or migrating is cumbersome.

These issues hinder compliance with federal standards and degrade security and performance, often necessitating dedicated Splunk expertise—something **NIS ARGUS™** mitigates by empowering end users directly.

4. NIS ARGUS™ Solution

NIS ARGUS™ integrates Databricks with Splunk, leveraging cloud-native design, a tiered data pipeline, and advanced AI/ML analytics to streamline performance and cut costs:

Key Architecture Components

- **Cloud-Native Storage:** Raw logs are stored affordably on platforms like Amazon S3, enabling cost-effective, long-term data retention.
- **Databricks Medallion Architecture:**
  - **Bronze Layer:** Raw data storage for traceability.
  - **Silver Layer:** Cleansed, filtered data ensuring accuracy.
  - **Gold Layer:** Aggregated data optimized for analytics and machine learning.
- **Optimized SIEM Data Flow:** High-value, enriched data from the Gold Layer is selectively forwarded to Splunk via streaming services (Kafka/Azure Event Hubs) or HTTP Event Collector (HEC), significantly lowering costs.
- **Advanced Analytics:** Databricks supports sophisticated analytics, anomaly detection, and ML models for predictive security insights and its platform is designed to evolve with rapidly changing technologies.
- **Splunk Integration:** The Databricks Add-on for Splunk allows seamless analytics directly within the Splunk interface, preserving existing workflows.
- **Scalable Implementation:** Designed for incremental adoption, minimizing operational disruption.

AI/ML Capabilities Enabled by NIS ARGUS™

NIS ARGUS™ augments this architecture with Databricks' automation, ML, and large-scale analytics capabilities.

Feature	Description	Pain Points Addressed
Log Quality Scoring	Leverage transformer-based NLP models (e.g., BERT) to assess and filter low-value or malformed log entries pre-ingestion.	Reduces ingestion volume and cost; improves data quality and SIEM performance.

Anomaly Detection at Scale	Apply models like Isolation Forests or DBSCAN to identify abnormal system or user behavior in streaming or batch logs.	Improves threat detection precision; reduces dependence on manual correlation rules.
LLM-Powered Root Cause Analysis	Integrate lightweight LLMs (e.g., Mistral, LLaMA 3) to summarize incidents, generate threat narratives, and suggest actions.	Reduces analyst workload and time-to-resolution; improves transparency and reporting.

Benefits of NIS ARGUS™

- **Cost Optimization:** Reduces Splunk data ingestion volume by up to 80%, significantly lowering licensing costs.
- **Enhanced Analytics and AI Integration:** Advanced anomaly detection, predictive modeling, and LLM-driven incident summarization within Databricks.
- **Operational Continuity:**
  - Existing SIEM workflows remain intact, ensuring smooth transition.
  - Integrates Databricks insights directly into the Splunk using the native add-on.
- **Improved Performance and Scalability:**
  - Accelerates threat detection and investigation through faster queries and optimized data pipelines.
  - Easily scales to handle growing log volumes and new data sources using cloud-native infrastructure.
- **Flexible Data Retention:** Immediate, actionable data retained in Splunk; historical data stored economically in cloud storage for extended retention and analytics.

Real-World Validation: NIS Databricks-Splunk Integration Proof of Concept (POC)

NIS validated **NIS ARGUS™** through an Azure-based POC integrating Databricks and Splunk:

- Reduced Splunk data ingestion by up to 80%.
- Enhanced query and dashboard performance.
- Enabled real-time monitoring, anomaly detection, and predictive analytics.
- Confirmed scalability, cost-effectiveness, and advanced analytical capabilities.

5. Conclusion

As Splunk licensing costs escalate, particularly following Cisco's acquisition, adopting cost-effective solutions becomes imperative. **NIS ARGUS™**, developed by NIS, offers a practical and scalable solution to reduce SIEM costs, enhance threat visibility, and meet federal cybersecurity mandates. Backed by our expertise in cloud computing, data analytics, and AI/ML, **NIS ARGUS™** empowers organizations to modernize their security operations while optimizing investments.

For more information, visit https://nw-its.com/capabilities/research-innovation.