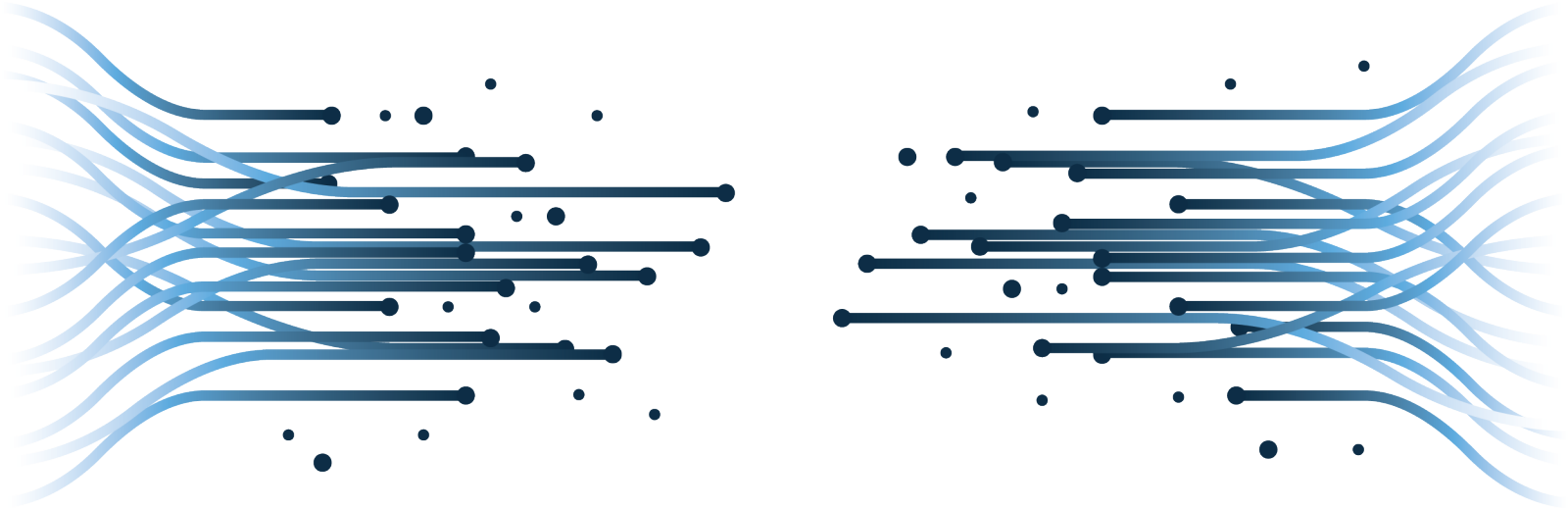
Optimizing SIEM Costs and Performance with Databricks Integration



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

Organizations increasingly depend on advanced analytics platforms for informed decision-making, yet rising licensing costs for Security Information and Event Management (SIEM) solutions like Splunk present significant challenges. Nationwide IT Services offers a cost-effective approach leveraging Databricks, enabling substantial cost savings, enhanced performance, and compliance with Executive Order 14028 and OMB M-21-31 cybersecurity standards. This paper outlines our innovative solution designed to maintain scalability, high performance, and regulatory compliance while optimizing expenses.

2. Background

Executive Order 14028, issued in May 2021, mandates federal agencies adopt advanced cybersecurity measures, including Zero Trust Architecture, multi-factor authentication, data encryption, and improved logging practices. Complementing this, OMB M-21-31 provides a phased maturity model for Event Logging (EL), progressing from basic (EL1) to advanced (EL3) capabilities within two years.

Splunk, a prevalent SIEM solution recently acquired by Cisco, faces potential licensing cost increases despite the promise of enhanced security solutions. To proactively address these financial challenges, our solution integrates Databricks, significantly reducing costs without compromising performance or security.

3. Platform and Data Challenges

Scaling SIEM platforms introduces critical challenges:

- **High Data Ingestion Costs**: SIEM licensing based on data volume makes storing exponentially growing log data financially unsustainable.
- **Processing Overhead:** Real-time data analysis requires substantial computational resources, creating performance bottlenecks.
- **Complex Real-Time Analytics:** Large, unfiltered data volumes complicate timely threat detection.
- **Data Quality Management:** Irrelevant data inflates costs, necessitating effective filtering.
- **Limited Advanced Analytics:** Traditional SIEM tools struggle with advanced analytics and machine learning (ML) due to architectural constraints.
- **Reduced Data Retention:** To cut costs, organizations may limit data retention, compromising critical historical insights and compliance requirements.

These issues hinder compliance with federal standards, impacting security and performance.

4. Proposed Solution Overview

Our architecture combines Databricks, AI capabilities, cloud-native storage, and targeted content filtering, reducing ingestion costs and enhancing analytics:

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 machine learning models for predictive security insights.
- **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.

Benefits

- **Cost Efficiency:**
  - Reduces Splunk data ingestion volume by up to 90%.
  - Efficient management of retention policies without compliance risk.
- **Enhanced Analytics Capabilities:**
  - Advanced trend analysis, predictive analytics, and improved threat detection powered by Databricks.
- **Operational Continuity:**
  - Existing SIEM workflows remain intact, ensuring smooth transition.
- **Performance Optimization:**
  - Faster query performance and scalable data management.
- **Flexible Data Retention:**
  - Immediate, actionable data retained in Splunk; historical data stored economically in cloud storage for extended retention and analytics.

Real-World Validation: Nationwide IT Services POC

Nationwide IT Services validated the architecture with an Azure-based pilot integrating Splunk and Databricks:

- Reduced Splunk data ingestion by **90%**.
- 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. Nationwide IT Services' innovative Databricks-based architecture provides an effective solution to reduce costs, enhance analytics, and maintain compliance with stringent federal cybersecurity standards. Our expert team delivers cutting-edge services in cloud computing, cybersecurity, data analytics, and AI/ML to help organizations optimize technology investments, ensuring sustained performance, security, and growth.

For more information, visit www.nw-its.com.