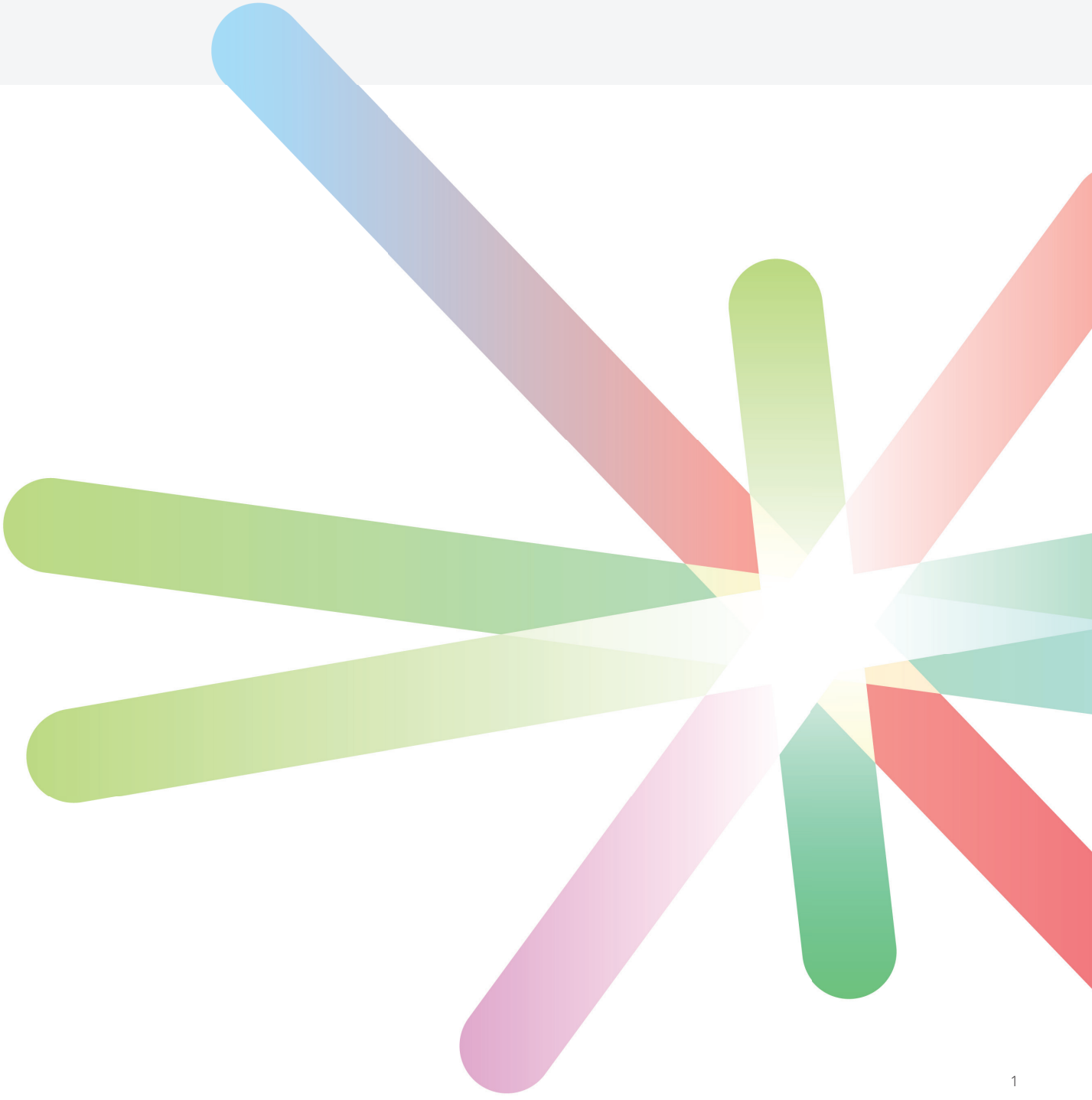
Deloitte. | databricks

ELEVATING PUBLIC SECTOR CYBER DEFENSE

CONSIDERATIONS FOR CHANGE AND THE CASE FOR A CYBER DATA PLATFORM

CONTENTS



EXECUTIVE SUMMARY

Public sector organizations are facing an unprecedented convergence of cybersecurity threats, regulatory demands, and operational complexity.

To address this, Deloitte and Databricks bring together the power of the Databricks Data Intelligence Platform with Deloitte's public sector experience, demonstrated implementation frameworks, and regulatory insight. Together, we enable government agencies to modernize security operations, realize the potential value of unified cyber data, and anticipate evolving adversaries.

INTRODUCTION

In the era of rapidly expanding data volumes and heightened regulatory demands, government entities aspire to anticipate persistent and evolving cybersecurity threats. But several challenges stand in the way:

DATA OVERLOAD

Managing significant data volumes from diverse sources

COST EFFICIENCY

Optimizing data usage while reducing costs

FOCUSED VISUALIZATION

Enhancing situational awareness for cyber operations and Security Operations Centers (SOCs)

ADVANCING CAPABILITY

Transitioning teams from reactive alerting to proactive threat hunting and advanced response with data use flexibility

REGULATORY AND OPERATIONAL LOGGING REQUIREMENTS

Meeting significant requirements from federal or sector-specific mandates, such as the Office of Management and Budget (OMB) M-21-31

In addition to these challenges, public sector entities now face intensifying pressures:

EXPONENTIAL INCREASE IN SENSITIVE DATA FROM DIGITAL AND CONSUMER BEHAVIORS

The proliferation of smart and connected devices (from homes and wearables to critical infrastructure) and the acceleration of digital services are producing huge volumes of sensitive data. Each new data stream expands the threat surface.

EXPANSION OF CRITICAL SYSTEMS AND THE ATTACK SURFACE

With more satellite networks, remote work connections, and interconnected ground systems, the number of access points for potential attack continues to grow.

ESCALATING SOPHISTICATION AND SCALE OF ATTACKS

Generative AI has empowered cyber adversaries to be more effective and dangerous by significantly increasing their productivity, sophistication, and ability to scale attacks, requiring agencies to go beyond basic perimeter defenses and adapt to continuously evolving threats.

GEOPOLITICAL AND SUPPLY CHAIN RISKS

Heightened geopolitical tensions and reliance on global supply chains for fundamental technology components has introduced new uncertainties and vulnerabilities into the security equation.

Addressing these pressures requires more than technology upgrades—agencies need a cohesive data strategy and end-to-end operationalization. The Databricks platform brings scalability and intelligence, while Deloitte enables fit-for-mission transformation, compliance, and results.

THE NEXT FRONTIER:

A JOINT DELOITTE AND DATABRICKS SOLUTION FOR CYBER DATA MANAGEMENT

The next frontier for broad cyber data management is the development of a cyber data platform, built on advanced lakehouse architectures that move beyond traditional Security Information Event Monitoring (SIEM) tools to integrate advanced analytics, anomaly detection, and artificial intelligence (AI) and machine learning (ML) insights.

Databricks provides the industry-leading open lakehouse platform—breaking down silos, delivering scalable analytics, and supporting advanced detection in real time. Deloitte brings experience implementing, optimizing, and aligning the Databricks platform with government requirements, regulatory mandates, and agency missions.

> SIEM and Security Orchestration, Automation, and Response (SOAR) tools remain important to a broad cybersecurity platform and serve as an integral component to enhancing security posture.

Public entities can address cyber defense challenges while remaining compliant by first collecting, aggregating, and visualizing security information. Following data brokering, the next-generation cyber data platform architecture unites security, analytics, and compliance capabilities with a data-centric approach.

Deloitte leverages the Databricks platform to move clients toward an integrated, scalable environment that meets advanced detection, compliance, and operational goals—securing and democratizing data at scale. Clients maximize value from their lakehouse investment by decoupling ingestion and storage from their legacy SIEM solutions and point tools, using a demonstrated approach to unlock the full potential of their investments and industry-leading commercial-off-the-shelf (COTS) technologies.

Based on our experience with complex environments and helping public sector clients, we have identified three elements to establishing a cyber data platform:

01

DATA MANAGEMENT AND LAKEHOUSE ARCHITECTURE

Collect, retain, and aggregate data in a cost-efficient, scalable, and compliant way.

02

CONTENT DEVELOPMENT AND ADVANCED USE CASES

Utilize and visualize data for traditional SIEM alerting, business operations, and advanced use cases including for anomaly detection, user and entity behavior analytics, network analysis, and more.

03

ANALYTICS AND AI

Leverage lakehouse architecture with integrated AI/ML to enable proactive monitoring, deeper investigation, and rapid response.

01 DATA MANAGEMENT AND LAKEHOUSE ARCHITECTURE:

ADDRESSING CYBER DATA PLATFORM FOUNDATIONS

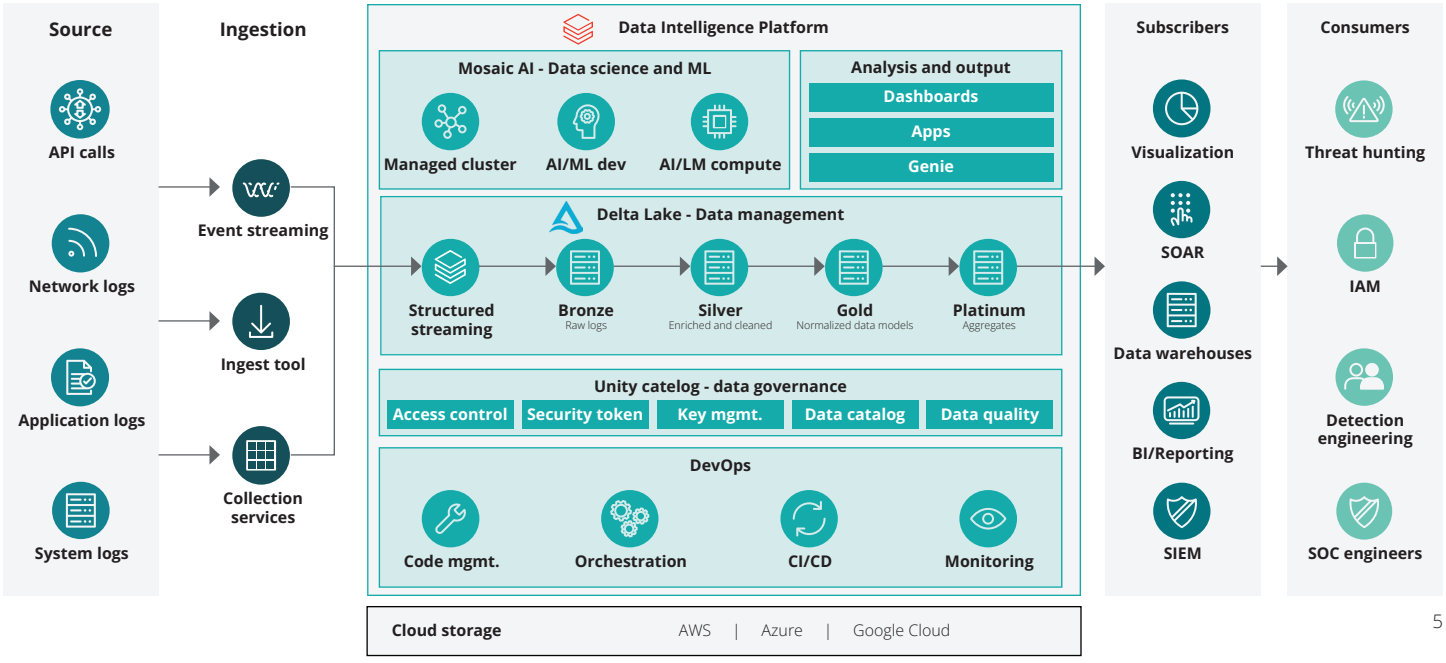
The effectiveness of cyber defense programs often hinges on visibility, high-quality data, correlation, and enrichment. Regulations on data logging and storage—alongside the expansive nature of rich data sources—compel agencies to assess their data management strategies while considering licensing, computing, and storage costs.

Lakehouse architecture such as the Databricks Data Intelligence Platform blends the scalability and flexibility of data lakes with the performance and governance of data warehouses. This enables cyber security teams to integrate data from diverse networks and systems—including stand-alone systems not directly connected to the SIEM—while maintaining compliance, real-time monitoring, and network capacity.

A multi-cloud, modern lakehouse architecture unifies cyber-relevant data—raw logs to analytics—on a single platform, breaking down organizational silos and simplifying ingestion, storage, and analytics. With automated schema enforcement, real-time integration, and elastic scalability, agencies can quickly adapt to mission demands, new technologies, and changing regulations while streamlining governance and reducing data duplication.

This centralized, AI/ML-ready platform democratizes access, enabling security, compliance, and analytics teams to draw from a single trusted source. As a result, investigations are accelerated, situational awareness is improved, and correlation of threats is enhanced, with flexible support for emerging data types and rapid deployment of advanced analytic models and automation.

DATABRICKS SECURITY LAKEHOUSE ARCHITECTURE



RECOMMENDATIONS

REEVALUATE CYBER SYSTEM ARCHITECTURE AND DATA MANAGEMENT

Assess existing architecture for effectiveness in achieving desired network activity visibility, analytic richness, and compliance adherence.

CONSIDERATIONS

- Integrating diverse data sources—legacy systems, cloud-native assets, IoT (Internet of Things), zero trust initiatives, remote workforce endpoints—presents new challenges. Measure data ingestion, aggregation, and maintenance against completeness, coverage, and compliance principles.

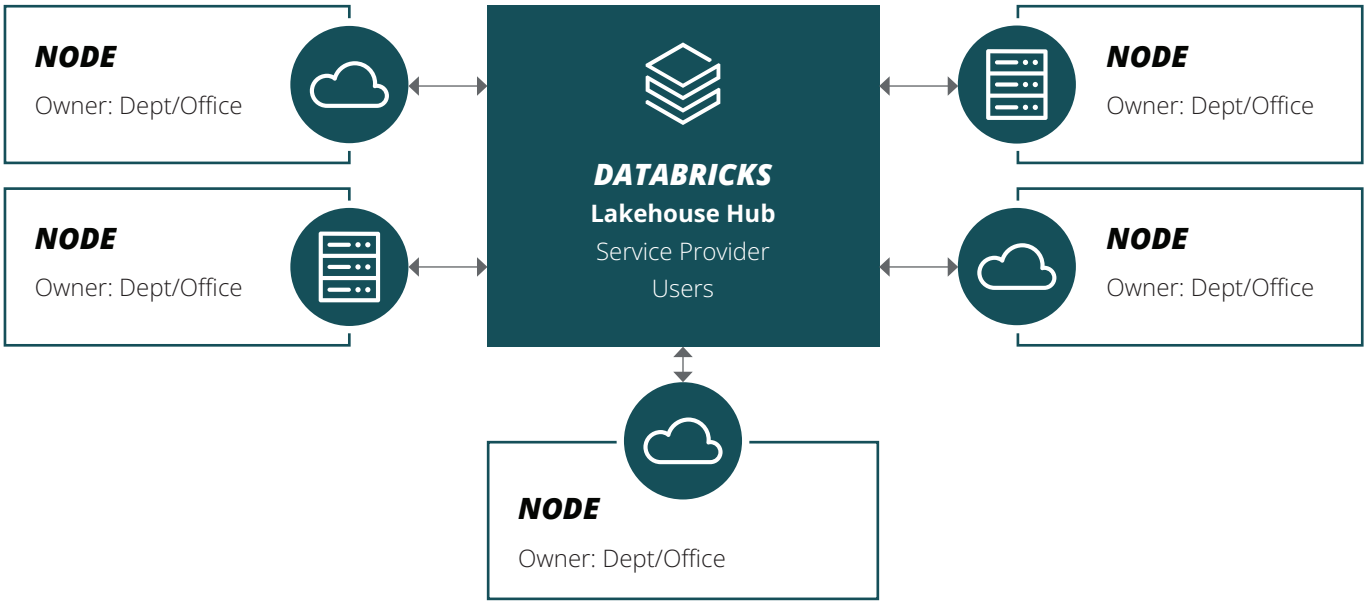
DEVELOP A FLEXIBLE LOG INGESTION APPROACH

Identify sources and appropriate methods (API, stream, log forwarders, third-party connectors) for integrating into the lakehouse and SIEM

CONSIDERATIONS

- Ingestion methods must support variety, including forwarders, APIs, syslog, and more and support multiple destinations and schemas.
- Environmental variables, endpoint and platform locations, and resource availability all shape real-time monitoring strategies.

A flexible collection approach is essential for purpose-built systems and the wide variety of devices used in complex enterprises, including mobile, IoT, and custom appliances. This avoids duplicate collection, enables full coverage, and enables centralized log aggregation, analytics, and reporting.



HELP MEET OPERATIONAL LOGGING AND COMPLIANCE REQUIREMENTS

Thoroughly assess collection capability, data types (malware, PII [Personally Identifiable Information], sensitive information) and anticipated data volume to meet mandates (OMB, HIPAA, FISMA, PCI, etc.)

CONSIDERATIONS

Data management at the platform level should ensure:

- **Collection**
  Data is gathered from each domain/enclave.
- **Completeness**
  Collect and store all necessary log elements centrally on the platform.
- **Coverage**
  Include all device types and sources contributing required logs.
- **Privacy**
  Integrate capability required to remain compliant with regulatory standards.

Explicit platform-level governance should guide selection and reporting of required data elements, while also reducing uncertainty and accelerating compliance.

DEVELOP A COST-EFFECTIVE AND EFFICIENT LICENSING AND STORAGE STRATEGY

Reassess how platform architecture, lakehouse capabilities, and cloud options impact data storage and licensing costs.

CONSIDERATIONS

- **The right storage/ retention model**
  Leverage a data tiering strategy to accommodate long-term incident investigations and threat hunting, balancing operational and 'cold'/archival storage with cost efficiency.
- **The ideal license model**
  Platform providers offer different pricing based on ingestion, compute and analytics.
- **Data transfer costs**
  Consider the total cost of data transfer, including ingress and egress, as well as storage, compute, and other total cost of ownership (TCO) factors.
- **Alternative retention model**
  Ingest logs into a central data lakehouse; selectively surface for SIEM or analytics based on use case or threat priority. The lakehouse capability allows direct search, advanced filtering, and retention flexibility, helping mitigate heavy licensing and processing requirements.

Modern lakehouse-based cyber data platforms for the public sector also provide:

- **Democratized access**
  Empowering interdisciplinary teams to work from a unified, trusted, and fully-governed data foundation.
- **AI/ML enablement**
  Enabling rapid, inline application of advanced analytics for faster threat detection and response.
- **Cost and operational efficiency**
  Reducing TCO through managed storage tiers, streamlined analytics, and right-sized infrastructure for maximum data ingest and search concurrency.
- **Future-ready flexibility**
  Supporting new use cases and data types with minimal operational disruption.

Together, a Databricks lakehouse and Deloitte delivery model can help agencies consolidate cyber data, improve outcomes and evolve their information assets.

02 CONTENT DEVELOPMENT AND ADVANCED USE CASES

POWERING PROACTIVE CYBER DEFENSE

For the public sector to pivot to proactive cyber defense, the cyber data platform should enable the foundational security content driving identification, prioritization, and response to close down the mean time to detect, respond, and resolve threat actor activity. The requirements go well beyond traditional log monitoring and SIEM alerting and should also address the evolving threat landscape.

ANOMALY DETECTION

Spotting novel irregular patterns that indicate insider threats or adversary behaviors. For example, automated threat investigation workflows—deployed through the Deloitte-Databricks solution—can now aggregate telemetry data from endpoints, networks, cloud sources, and applications. When an unusual login is detected at an atypical hour or from an unexpected geo-location, real-time correlation across user sessions, device health, and threat intelligence feeds can enable rapid escalation and triage, reducing response time from hours to minutes.

USER AND ENTITY BEHAVIOR ANALYTICS (UEBA)

Identifying privilege misuse, lateral movement, or cross-domain anomalies. Built-in UEBA capabilities profile typical user/device behaviors. In one public sector scenario, profiling flagged a privileged user accessing sensitive assets outside of their historical pattern. Such deviations are detected automatically, accelerating investigation and risk mitigation before misuse can occur.

NETWORK ANALYSIS AND INVENTORY

Understanding traffic flows, asset mapping, and device roles for threat correlation.

AUTOMATED THREAT INTELLIGENCE FUSION

Continually incorporate emerging adversary tactics into monitoring and analytics logic. Integration with external cyber threat intelligence enables the platform to receive and process new indicators of compromise (IOCs), rapidly updating detection content and triggering automated incident response playbooks. For instance, when a critical new malware hash is published, response playbooks can auto-isolate affected assets or block malicious activity in near real time, minimizing risk exposure.

INTEGRATING SOAR CAPABILITIES

Transform repeatable playbooks into orchestrated automated actions—dramatically reducing detection-to-response time.

Deployment example

- **Phishing triage and remediation**
  Databricks ingests user-reported emails and mail telemetry, extracts URLs/attachments, enriches indicators, and correlates to prior campaigns to generate a disposition and risk score. SOAR consumes the output to quarantine the original (and optionally "similar") messages, block malicious senders/domains/URLs, and open an incident with a packaged evidence set for review and audit.
- **Automated containment of malicious indicators**
  Databricks correlates detections across multiple sources to produce high-confidence IP/domain/hash indicators with context such as prevalence, affected assets, and confidence level. SOAR validates against allowlists and policy thresholds, then pushes blocks to the appropriate device APIs, confirms changes were applied, and records actions and exceptions in the case record.
- **Privileged access anomaly response**
  Databricks correlates identity and admin activity (sign-ins, privilege changes, device posture, and high-risk actions) to detect anomalies and explain what drove the alert. SOAR orchestrates the response workflow—step-up authentication, temporary access restrictions or suspension where policy permits, credential rotation workflows as needed, and routing to an approver or incident lead with a concise summary and recommended next steps.

A STRATEGIC IMPERATIVE

Organizations should carefully identify cases that will deliver the highest impact for their specific environment, risk landscape, and regulatory context while minimizing strain on overloaded operators. After identifying the right use cases, organizations should prioritize, engineer, and operationalize them. This approach transforms basic monitoring into a dynamic, proactive defense—one that can be machine automated to relieve operators of repetitive, low-level tasks and allow them up to focus on more complex, human-worthy challenges.

RECOMMENDATIONS

COLLABORATIVE USE CASE DISCOVERY

Work closely with stakeholders across threat management, incident response, threat intelligence, and engineering to define and refine use case libraries tailored to organizational needs.

SELECT CONSIDERATIONS

- Engage business and technical stakeholders to capture a full set of risks and regulatory drivers.
- Prioritize use cases that address high-impact threats, compliance gaps, or mission-critical assets.

SYSTEMATIC PRIORITIZATION

Group and roll out use cases by risk, regulatory requirements, and operational urgency. Prevent false positives and alert fatigue by iteratively tuning detection logic.

SELECT CONSIDERATIONS

- Establish clear criteria for determining use case criticality, such as potential impact and likelihood.
- Regularly review and tune detection logic to reduce false positives and optimize SOC workload.

RAPID CONTENT EVOLUTION

Integrate new data sources, such as IoT or cloud, and threat intelligence into use-case development and detection patterns to continuously advance analytic capabilities as threats and technologies change.

SELECT CONSIDERATIONS

- Maintain a flexible, modular detection framework to simplify onboarding of new data types.
- Collaborate with threat intelligence teams to adapt content based on emerging adversary tactics and IOCs.

USER DATA TO THE DEFENDERS' ADVANTAGE

Incorporate enrichment and contextualization to extend the quality of alerting, tickets, and bring broad decision-making and rapid response to the operators' fingertips.

SELECT CONSIDERATIONS

- Leverage contextual data—such as user behavior, device health, and asset criticality—to prioritize and enrich alerts.
- Develop automated enrichment workflows to provide analysts with relevant context at triage.

AUTOMATION

Identify and eliminate repetitive, low-level tasks and use playbooks and a full Configuration Management Database (CMBD) to drive automated remediation based on institutional risk tolerance and asset criticality.

SELECT CONSIDERATIONS

- Define clear triggers for automation to minimize manual intervention and escalate only complex cases.
- Regularly update playbooks to reflect new attack methods and business process changes.
- Maintain a current, accurate CMDB to increase automation value and accuracy.

03 ANALYTICS AND AI

DRIVING NEXT-GENERATION CYBER OPERATIONS

A lakehouse-based cyber data platform fundamentally elevates how organizations achieve advanced detection and analytics. By harnessing the Databricks lakehouse platform—supported by Deloitte's deep public sector and AI implementation experience—agencies can accelerate the deployment of future-proof cyber defense capabilities at scale, all while maintaining regulatory rigor and operational resilience.

Unlike legacy security architectures that rely on disjointed tools and basic statistical alerting, a unified lakehouse approach makes it possible to fully integrate, automate, and elevate core cyber operations via scalable, intelligent technology and probability such as:

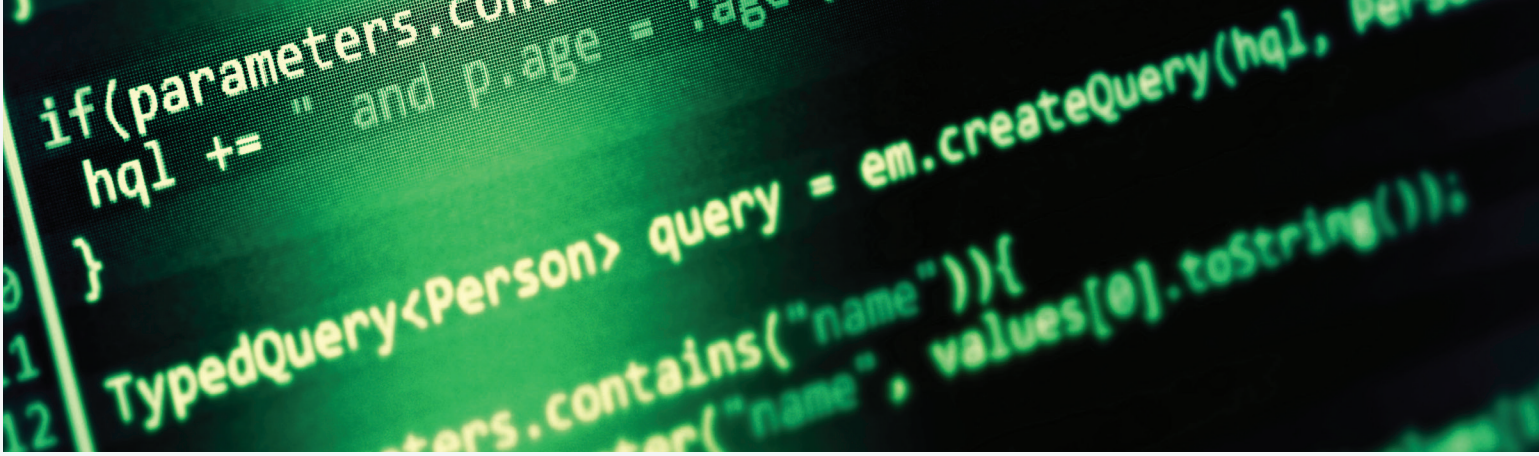
RECOMMENDATIONS

HIGH-FIDELITY DATA INGESTION AND REAL-TIME ANALYTICS

Ingestion of high-quality telemetry and contextual data—from endpoints, cloud environments, applications, and networks—lays a sound foundation for security analytics. Modern lakehouse platforms provide instant, unified access to this data, enabling security operations teams to investigate, correlate, and escalate incidents with unprecedented speed and clarity.

CONSIDERATIONS

- Confirm ingestion pipelines accommodate diverse formats, sources, and velocity (APIs, log forwarders, streams).
- Enable instant analysis for large-scale data, supporting ad hoc queries and long-term forensic investigation.
- Prioritize scalable architectures to handle surges in volume during active threats or compliance audits.



SHIFT FROM DETERMINISTIC TO PROBABILISTIC

Static, rule-based detection often yields noisy alerts and misses emerging tactics. Lakehouse architectures support probabilistic and risk-based approaches, tuning detections to context and likelihood rather than rigid thresholds—ultimately lowering false positives and elevating true threat detection.

CONSIDERATIONS

- Adopt statistical models (e.g., supervised/unsupervised learning) to infer normal vs. abnormal behavior.
- Incorporate uncertainty and risk scoring to drive prioritization and escalation decisions.
- Regularly calibrate detection models to adapt to evolving business and threat environments.

EMBED AI, MACHINE LEARNING AND GENERATIVE AI (GENAI)

Integrating AI and machine learning enables systems to uncover sophisticated, previously unknown behavioral patterns or zero-day techniques across enormous, unified datasets. In parallel, GenAI can act as an analyst "copilot" for incident response and threat hunting, accelerating how teams analyze data, summarize cases, and translate detections into operational playbooks. Together, machine-driven signature discovery and GenAI-assisted investigation reduce manual effort, improve consistency, and surface novel threats faster.

CONSIDERATIONS

- Curate high-quality training data from across organizational silos.
- Enable governed GenAI workflows such as Databricks AI/BI Genie with natural language interrogation, automated case summarization, and hypothesis generation.
- Regularly retrain and validate models to capture new patterns, reduce false positives, and minimize model drift.
- Use GenAI to accelerate detection logic and SOAR playbook generation (templates, mappings, test cases), with human review and version control.
- Automate threat signature creation and update workflows for faster operationalization.

PLATFORM NATIVE ADAPTABILITY

A flexible, cloud-native lakehouse platform supports onboarding of new asset types—such as IoT sensors, remote endpoints, and evolving SaaS solutions—without architectural rework. This adaptability is essential for sustaining long-term cyber resilience amid rapid technology shifts.

CONSIDERATIONS

- Use open, extensible connectors to incorporate diverse device and application telemetry.
- Support flexible data models that recognize new event schemas as the environment changes.
- Enable centralized governance and access management for both legacy and next-generation assets.

ACHIEVE MACHINE TO MACHINE SPEEDS

By orchestrating machine-driven investigation, enrichment, and response, organizations can react and adapt at "machine speed"—outpacing adversaries who rapidly weaponize automation and AI. End-to-end agentic workflows minimize manual bottlenecks and enable continuous defensive adaptation.

CONSIDERATIONS

- Build automated playbooks for detection, investigation, and response activities.
- Integrate sensor-based feedback loops to adjust detection thresholds in real time.
- Continuously benchmark workflow speed and adaptability to evolving adversary tactics.

NEXT STEPS TO OPTIMIZE YOUR CYBER DATA SYSTEMS:

01

Integration with cyber threat intelligence (CTI)
Integrate CTI to enhance visibility by incorporating indicators of compromise (IOC) into SIEM monitoring. Databricks can ingest and analyze threat intelligence feeds, enriching SIEM content and supporting dynamic risk assessment.

02

Champion responsible analytics
Govern use of ML and automation with compliance-aligned policies to maintain transparency, trust, and control.

03

Continual training and optimization
Enable teams to maximize value from platform features, analytic workflows, and automation capabilities to capitalize on new releases and data sources.

WHY DELOITTE?

As regulated industries advance their cyber defense strategies, Deloitte provides broad experience and demonstrated capabilities to support complex transformation initiatives.

Architect, implement, and augment lakehouse-based cyber data platforms aligned with rigorous regulatory requirements and organizational needs.

Integrate and modernize SIEM, SOAR, UEBA, and advanced analytics to enhance protection and improve operational efficiency.

Accelerate compliance initiatives and future-ready technology infrastructure through established methodologies, proprietary tools, and recognized practices.

Offer ongoing support, training, and strategic insight as cyber threats and the technology landscape evolve.

Deloitte can serve as your Databricks managed service provider (MSP) and/or managed security service provider (MSSP), operating and securing the lakehouse either within an agency's Databricks environment or within Deloitte's compliant managed services environment – where permitted. This approach provides a reliable, audit-ready platform and accelerates cyber outcomes through end-to-end operations, security acclimatization, identity and access management, continuous monitoring, and cost efficiency, while enabling telemetry-driven detection, threat hunting, and investigation supported by defined service-level agreements (SLAs), governance, and executive reporting.

NEXT STEPS

FROM VISION TO CYBER RESILIENCE

Enhancing cyber data management, detection, and response capabilities is imperative for regulated industries facing escalating threats and compliance demands. The move to modern cyber data platforms—and especially lakehouse architecture—empowers organizations to achieve superior visibility, proactive defense, and broad compliance.

While we've focused on foundational imperatives and recommended strategies, it is important to note that Deloitte and Databricks offer broader support for cyber data optimization. Beyond platform architecture and advanced analytics, our teams can assist with cloud migration, data governance, advanced reporting, compliance automation, and more—helping organizations address evolving challenges and realize greater value from their cyber data assets.

A unified, end-to-end approach harmonizes tool choices, leverages advanced analytics, and supports real-time actionable insight. Deloitte stands ready to work together to help agencies build a next-generation cyber data platform with the proven capabilities and strategic vision to realize your organization's cyber goals.

Schedule a consultation with our team to discover how to maximize value from your cyber data assets.

AUTHORS

Bob Cheripka
bcheripka@deloitte.com
Tech Fellow, GPS Cyber Risk
Deloitte Advisory LLP

Scott Reide
jscottriede@deloitte.com
GPS Specialist Leader
Deloitte & Touche LLP

Kent Myer
kentmeyer@deloitte.com
GPS Cyber Detect & Respond, Managing Director
Deloitte & Touche LLP

Mark Lopez
marklopez@deloitte.com
GPS AI & Data, Specialist Master
Deloitte Consulting LLP

Joe Nehila
jnehila@deloitte.com
GPS Cyber Detect & Respond, Manager
Deloitte & Touche LLP

Emily Cole
emcole@deloitte.com
GPS Databricks Alliance Manager
Deloitte Consulting LLP