

Beyond SIEM: Building a Detection-First Security Data Architecture



How Integrated Data Engineering and Detection Posture Management Unlocks Transformational SOC Outcomes

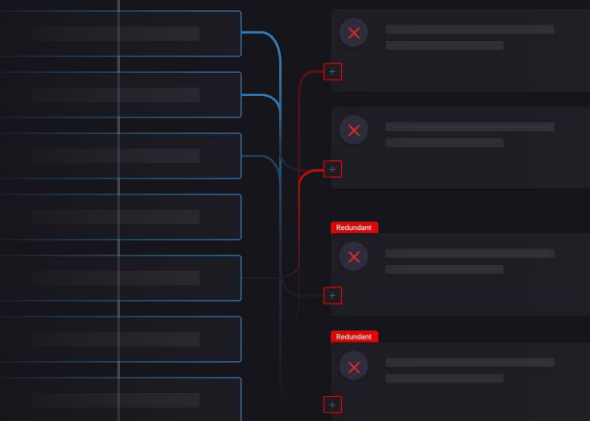
Security Information and Event Management (SIEM) platforms have long served as the operational backbone of the Security Operations Center (SOC). For years, organizations built their security telemetry processes around legacy SIEM architectures: collect everything, centralize it all in one place and normalize schemas, then build detections on top of it all.

But today's environments look nothing like what early SIEM architectures were designed for. Multi-cloud deployments, SaaS sprawl, containerized workloads, API-driven infrastructure, and identity-centric attack paths have dramatically expanded telemetry volume AND complexity. Legacy SIEM data infrastructure and licensing models lead to an excruciating choice: pay exorbitant licensing fees to keep all their data, OR prioritize cost savings while losing visibility and coverage.

This tradeoff has driven the shift from legacy SIEM data models to modular security data pipelines and data lakes that promise flexibility, cost control, and scalability. But they also introduce a critical question: how do you translate more data, in more places, into better security outcomes?

The answer to that question lies in the intersection of *security data engineering* and *detection engineering*. Without embedding detection engineering and posture management processes into data engineering workflows, modern security data platforms risk becoming cheaper storage systems rather than engines of effective threat detection and risk mitigation.

Why Legacy SIEM Data Architectures Are Under Strain



SIEM architectures were designed around a simple model: ingest, normalize, and store as much IT and security telemetry as possible, just in case you'll need them later. Then build detection rules inside the SIEM to generate alerts for incident response and investigation. Unfortunately, this legacy SIEM model didn't anticipate (or care about) the explosion in log volumes from increasingly complex enterprise environments.

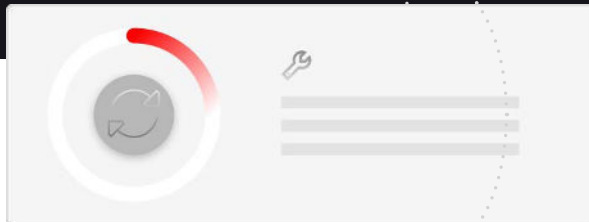
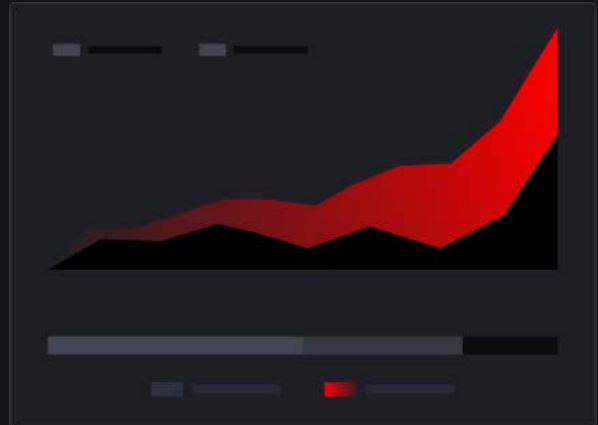
Challenges with Legacy SIEM Models

Over time, three structural problems emerged with legacy SIEMs.

Runaway Data Costs

SIEM licensing models tied to ingestion volume have become one of the fastest-growing line items in security budgets. As organizations ingest cloud logs, identity telemetry, endpoint data, SaaS events, and custom application logs, costs scale quickly.

The natural response is to filter aggressively at ingest time. But filtering data to reduce costs often means sacrificing visibility and detection coverage, forcing tradeoffs between budget control and detection coverage.



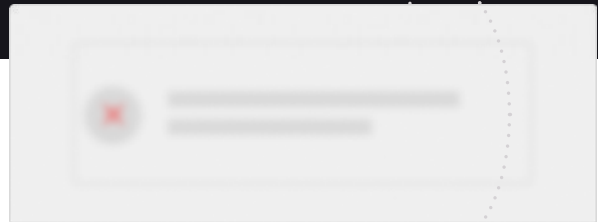
Operational & Maintenance Burden

Exploding IT infrastructure and attack surfaces have forced modern SOCs to evolve and prioritize data engineering as a critical function. Teams now spend enormous effort parsing logs, maintaining pipelines, handling schema drift, and troubleshooting broken integrations. And the more log sources you add, the more brittle your detection stack becomes, putting enormous pressure on detection engineers to build and maintain detections for all this telemetry.

A key finding from our [2025 CardinalOps State of SIEM Detection Risk Report](#) sheds light on the disconnect between data ingestion and effective detection coverage:

On average, organizations ingest enough telemetry to potentially cover 90% of MITRE ATT&CK techniques, but only have detections in place for 20% of those techniques.

In other words, the problem is rarely the lack of data. It's a failure to operationalize that data into reliable, high-quality detections. And with legacy SIEM licensing models, teams are not only missing threat coverage opportunities. They're effectively burning cash in the process.



Hidden Detection Failures

Even when detections exist, a nontrivial proportion of them break and fail silently. Log source changes, field renaming, parser updates, schema drift, dependency failures, and misconfigured scheduling parameters can all cause rules to stop firing—without anyone noticing. Our State of SIEM Detection Risk Report also included an alarming finding on this front:

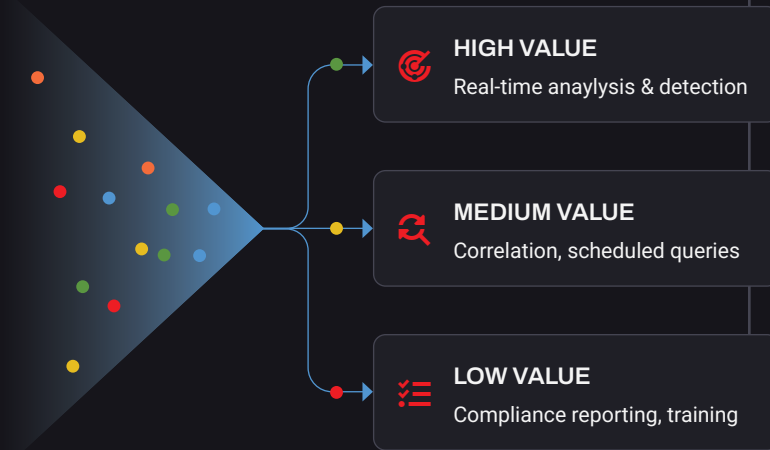
13% of the rules across enterprise SIEM environments were broken.

In other words, more than 1 in 8 of detections silently fail, potentially allowing a threat to go unnoticed and present significant risk in your environment. That means many SOCs have a dangerous illusion of coverage. Dashboards show green, but critical attacker behaviors go undetected.

Security Data Pipelines and Lakes: A Structural Shift to Modularity and Flexibility

To address these issues, many organizations are rearchitecting their data engineering processes around modular pipelines and data lakes. Instead of routing all telemetry directly into a monolithic SIEM for indexing and detection, modern architectures use a more surgical approach where pipelines:

- ✔ Ingest and route telemetry using contextual logic to filter out noise.
- ✔ Normalize and enrich data before it reaches analytics engines.
- ✔ Send high-value data to “hot” storage for real-time detection.
- ✔ Send medium-value data to “warm” storage for correlation and scheduled queries.
- ✔ Store low-value logs in cost-effective data lakes for compliance and LLM model training.



This model decouples data storage from detection logic and introduces greater flexibility in how data is processed and enables multiple downstream use cases. The benefits are significant:



Cost Optimization

organizations can route rarely queried data to inexpensive cold storage and skip resource-intensive data processing modules. This reduces security data costs while giving SOC teams the option to reheat and rehydrate the raw telemetry for future investigations.



Agility

modular building blocks means data infrastructure can be updated independently of detection logic. Teams can adopt new analytics engines, cloud-native tools, or AI-based platforms without rebuilding the entire structure from the ground up.



Data Governance

centralized normalization and schema management for high-value security data create more consistent, high-quality telemetry across the security stack for detection workflows. And since not all data is needed for real-time detection, modern pipelines include routing logic to identify lower-value data, skip normalization, and send it to low-cost storage.

These flexible architectures solve cost and scalability challenges, but they do not automatically improve detection outcomes. They can actually introduce new blind spots if detection engineering is not embedded into the data lifecycle itself.

More Data ≠ Better Detection Coverage

It's tempting to assume that if you fix the cost model and improve data accessibility, detection quality will follow naturally. In practice, the opposite can happen.

Expanding your visual field doesn't mean you're actually seeing what's important. So when data is distributed across pipelines, analytics tools, and storage destinations, **new risks emerge**:



It becomes less obvious which detections rely on which log sources.



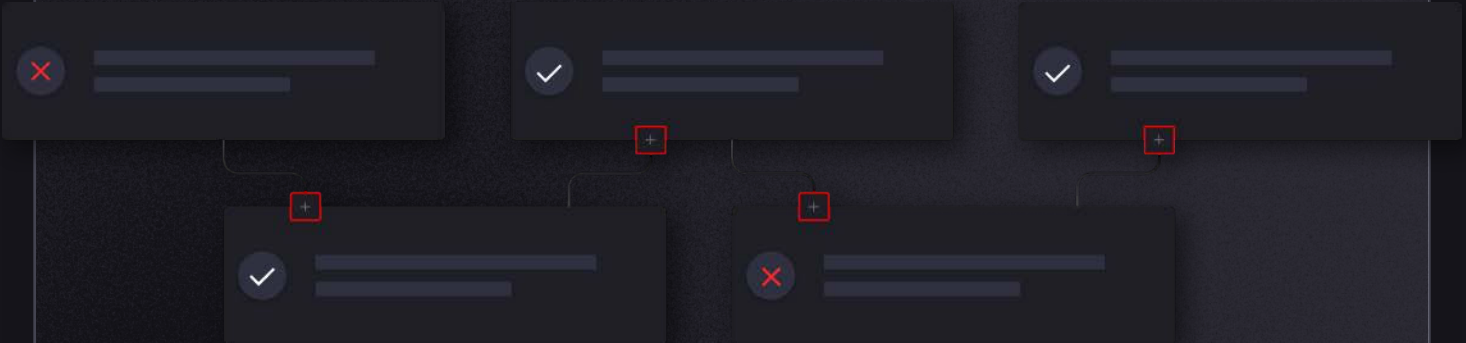
Schema drift in upstream pipelines can silently break downstream rules.



Data quality issues propagate across environments.



Detection logic fragments across platforms.



If detection engineering remains manual and reactive, SOC teams end up with the same core problem in a new architecture: abundant telemetry but inconsistent, unreliable coverage. Achieving a high level of detection maturity goes beyond just better data plumbing. It requires a deliberate intersection between the disciplines of data engineering and detection engineering focused on collaboration and continuous refinement.

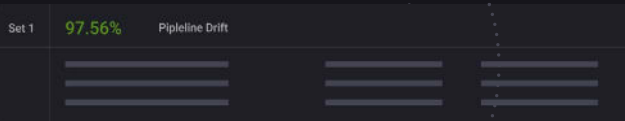
Detection Engineering as a Security Data Control Plane

In a modern SOC, detection engineering must evolve from a rule-writing task inside a SIEM to a control plane governing the full IT and security telemetry lifecycle. Integrating detection engineering automation and posture management with modern security data engineering processes helps provide that control plane in several key ways:



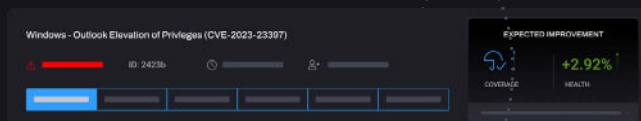
Mapping Coverage to Telemetry

Mapping detections to adversary techniques (e.g., MITRE ATT&CK) reveals the gap between potential and actual coverage. When telemetry exists but no detection leverages it, AI-powered detection engineering workflows surface high-priority coverage gaps and works to close them with new detections. Processes to create and maintain an inventory or similar document that outlines which detections rely on which log sources sets a strong foundation that the entire system stands on.



Monitoring Pipeline Drift

As new log sources are added, schemas are updated, and normalization logic is changed, data pipeline requirements naturally evolve. Embedding detection engineering into routine processes that continuously refine those requirements ensures that detection coverage remains strong while preventing noisy alerts from flooding response workflows. It also ensures that security data architecture remains resilient and cost-effective.



Continuously Validating Rule Health

Instead of assuming detections work, automated validation checks ensure that required fields exist, parsers function correctly, dependencies resolve properly, and rules fire as expected. This directly addresses the less silent failure problem with complex SIEM environments. It not only keeps your detection posture strong but ensures ROI on data engineering efforts to ingest and process security data.

RULE NAME	MITRE ATTACK	
	TACTIC	TECHNIQUE
1		
2		

Embedding Detection-as-Code

Introducing complex, nested logic into both data processing and detection engineering workflows makes shifting left imperative. For peak SOC performance, detection-as-code isn't a nice-to-have, it's essential. When detection logic is version-controlled, tested, and deployed through CI/CD pipelines, changes in data schemas or telemetry sources can be validated \ and rules can be tuned automatically, before downstream impacts on production systems.

The Best of Both Worlds: Cost AND Detection Posture Optimization

Security data pipelines and lakes are often justified on cost savings alone. That's a mistake. The real strategic opportunity lies in shifting from cost optimization to a broader view of detection posture optimization:

- ✓ Route low-value data away from high-cost platforms.
- ✓ Invest in telemetry that directly supports high-priority detections.
- ✓ Continuously measure coverage against adversary behaviors.
- ✓ Treat broken or noisy rules as measurable risk.
- ✓ Automate tuning to maintain high signal fidelity.

Without integrated detection engineering automation, organizations may reduce SIEM bills, but still lack reliable coverage of critical attacker techniques. With integrated detection posture management, however, modern data architectures become powerful enablers. Data lakes preserve long-term visibility without bankrupting the SOC. Pipelines ensure the right data reaches the right analytics engines, at the right time. Automated detection engineering ensures that telemetry translates into meaningful, high-confidence alerts.

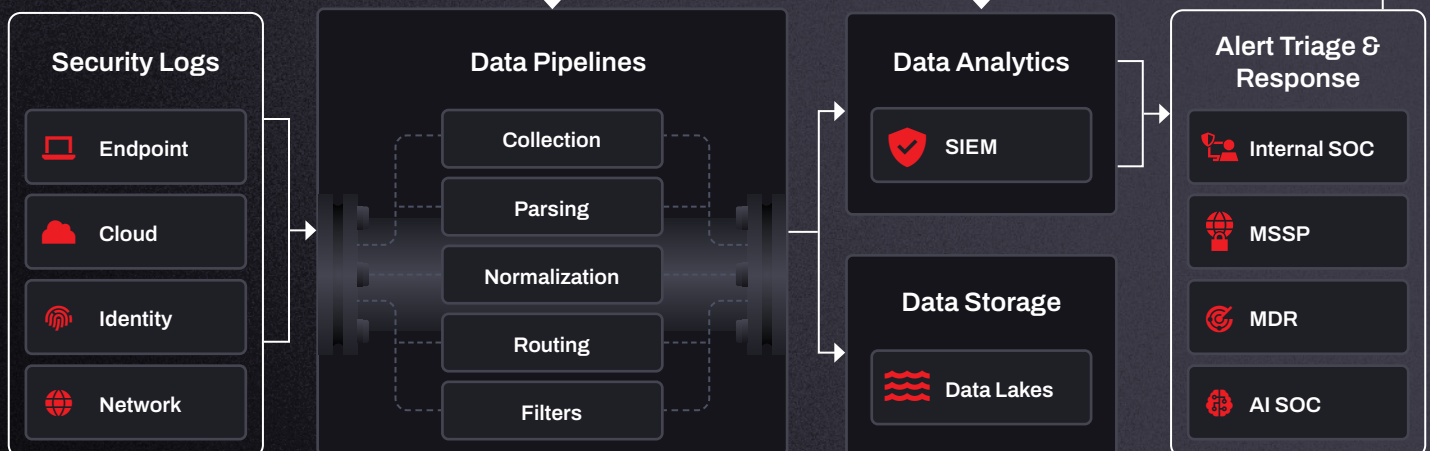
AI-Powered Detection Posture Management

Data Engineering

- ✓ Missing events & fields
- ✓ Parsing errors
- ✓ Cost savings from redaction & unnecessary log filtering

Detection Engineering

- ✓ New rules for coverage gaps
- ✓ Rule logic fixes for broken and noisy detections
- ✓ Alert fidelity insights





Looking Ahead

The transition from legacy SIEM-centric data infrastructure to modular security data platforms is accelerating. The economics and architectural realities of modern environments demand it. SOC leaders need to see this shift as a structural opportunity, not just an incremental cost-savings tactic.

Organizations that treat security data pipelines as only a cost control measure will likely achieve efficiency but miss out on significant security posture benefits. Embedding automated detection engineering into their data architecture provides something far more valuable: measurable, continuously improving threat coverage across a scalable, cost-effective platform.

The ultimate goal is more than just moving logs faster and paying less for them. It is to reliably detect the attacker behaviors that matter most.

Modern security data pipelines make that possible. Integrated detection engineering automation and posture management makes it real.

Interested in learning how CardinalOps can embed automated detection engineering and posture management into your modern security data architecture? **Let's chat.**

[Contact Us](#)