pwc

Data Pipeline Management: Building and Operating the Security Data Layer

Executive summary

Security operations centers are overwhelmed by rapidly growing volumes of telemetry from endpoints, networks, cloud services, identity systems, and IoT environments. While this data growth should, in theory, improve detection and response, most organizations are experiencing the opposite outcome: rising cost, growing complexity, and declining visibility [1].

Traditional SIEM platforms centralize security data but rely on ingestion-based pricing models that do not scale with modern telemetry growth. As data from cloud, endpoint, identity, and network sources accelerates, organizations are forced into trade-offs such as reducing logging, aggressively filtering events, or shortening retention periods. Each of these measures lowers cost, but also erodes detection coverage and investigative depth, increasing operational risk [1].

This creates a fundamental contradiction. Enterprises ingest enough telemetry to theoretically enable detection of more than 90% of MITRE ATT&CK techniques, yet most SIEM deployments operationalize detection logic for only ~21% of those techniques. The gap is not caused by a lack of data, but by the inability of existing architectures to transform raw telemetry into sustained, high-quality detection engineering [16].

A significant share of security budgets is now consumed by SIEM ingestion costs and overlapping tools, rather than investments that measurably improve detection quality or reduce attacker dwell time.

This white paper outlines a structural alternative based on security data pipeline management. Instead of routing all telemetry into a single SIEM, security data is filtered, enriched, and routed according to purpose. High-value signals are delivered to the SIEM for real-time detection and investigation, while data required for compliance or long-term analysis is stored in lower-cost, searchable data lakes or object storage. Telemetry with no security or regulatory value is removed or masked. This approach reduces SIEM cost pressure while preserving retention, analytical flexibility, and detection effectiveness.

Data pipeline management: the new solution to scale cyber defense

A security data pipeline is a control layer between data sources and destinations such as SIEMs, data lakes, and XDR platforms. It ingests telemetry, **applies filtering and enrichment, and routes data** in near real time.

As security data volumes grow, pipelines have become essential to maintain data quality and control costs. They ensure that only relevant, well-structured data is consumed, **directly impacting detection accuracy and operational effectiveness**—especially as analytics and AI rely on consistent, high-quality inputs.

While destination technologies may change, a managed data pipeline provides **stability, governance, and flexibility** across the security architecture.

Core capabilities of a security data pipeline include:

- **Ingestion and normalization**: Telemetry from diverse systems is normalized into consistent schemas so downstream tools receive structured data and correlation becomes feasible, **reducing brittle, tool-specific parsing layers** [3].
- **Filtering and volume control**: High-volume sources are deduplicated, aggregated, or filtered early to limit ingestion costs, protect SIEM performance, and keep analyst focus on higher-value signals [3].
- **Enrichment and lightweight correlation**: Events are enriched with asset context, location, vulnerability data, and threat intelligence, with basic correlation applied before data reaches analytic platforms to **reduce triage effort** [1].
- **Masking and encryption**: Sensitive fields can be masked, tokenized, or encrypted centrally, ensuring consistent privacy controls and supporting audit and traceability requirements.
- **Selective routing and tiering**: High-value signals are routed to SIEMs, while bulk or long-term data is sent to lower-cost storage where it remains searchable, allowing **detection and storage to scale independently** [4].
- **Federated search support**: Analysts can query data across multiple storage locations without re-ingestion, **reducing duplication and cost** while preserving investigative capability [4].

Business benefits of data pipeline management (1/2)

By Data pipeline management delivers value beyond technical optimization by directly addressing cost, detection effectiveness, compliance, and long-term resilience. By controlling how security telemetry is ingested, enriched, and routed, organizations can reduce operational friction while improving the quality and timeliness of security outcomes. This shifts security data from a cost driver into a managed asset that supports both risk reduction and business agility.

Figure 1: Overview of business benefits for data pipeline management

From a business perspective, data pipeline management delivers:

- Cost optimization and predictable scaling: Filtering and routing data before it reaches the SIEM reduces ingestion and storage costs. High-volume raw telemetry can be diverted to lower-cost, searchable storage without sacrificing investigative capability, providing immediate budget relief as SIEM costs rise with data volume [3][4].
- Improved detection quality and analyst productivity: Noise reduction and enrichment ensure that only relevant, context-rich events reach analytic platforms. Pre-processing in the pipeline shortens detection and response cycles and allows SIEMs to focus on analysis rather than data handling [4].
- Compliance and audit readiness: Pipelines enforce retention policies, apply masking to sensitive fields, and preserve complete audit trails. Searchable archived data supports faster responses to audits and investigations and helps maintain compliance across major regulatory frameworks [1].

Business benefits of data pipeline management (2/2)

- Agility and future-proofing: By decoupling data ingestion from analytics, pipelines allow organizations to change SIEMs or adopt new analytics platforms without re-engineering data flows or losing historical data. The same structured data layer also supports emerging AI-driven use cases [1][3].
- Business alignment and risk reduction: Reliable, high-quality telemetry improves early detection and containment while providing evidence of due diligence. At the same time, reduced ingestion overhead frees analyst capacity for higher-value security activities.
- Faster detection through shift-left data processing: Applying enrichment and correlation directly in the data stream reduces delays introduced by traditional SIEM processing, limiting attacker dwell time and improving MTTD and MTTR [5].
- Improved collaboration and development efficiency: Detecting issues earlier in the data flow reduces remediation cost and rework. Integrating security controls upstream improves coordination between teams, raises quality, and reduces downstream fixes [6][7].
- Innovation without lock-in: Handling data quality, governance, and enrichment upstream reduces duplication and inconsistency while lowering processing cost. This supports faster decision-making, self-service analytics, and AI-driven innovation without increasing platform dependency or operational overhead [8].

Architectural blueprint for a modern security data pipeline

A modern security data pipeline is a modular architecture that securely ingests, normalizes, filters, and routes security telemetry based on purpose and cost. Strong governance, federated analytics, and centralized secret management ensure efficiency, compliance, and adaptability as data volumes and technologies evolve.

The diagram illustrates data flowing from sources through ingestion, normalization, filtering and enrichment, then splitting into real-time detection paths and cost-efficient archival paths. Cross-cutting controls such as encryption, identity and secret management, metadata integrity and governance are applied throughout, highlighting modularity and resilience.

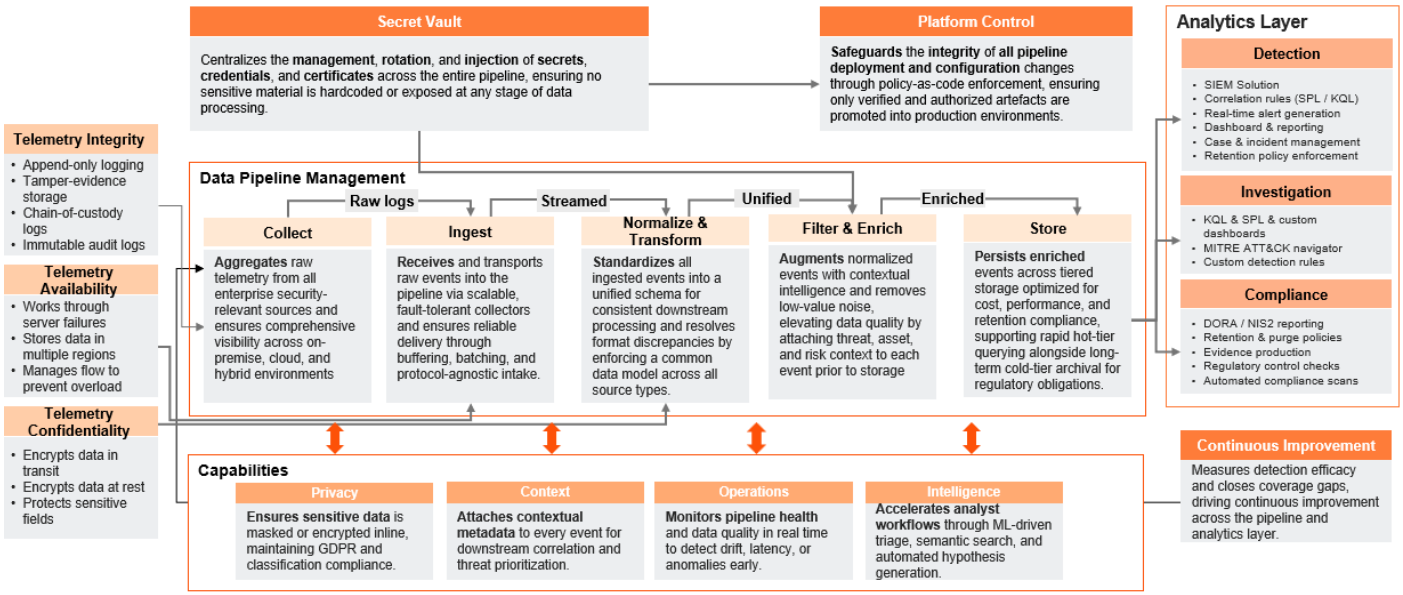


Figure 2: Modern security data pipeline architecture with extended use cases and secret management integration

Extended security use cases and control requirements

Security data pipelines reduce cost and improve detection, but they also sit on the critical path of security operations. Once telemetry is filtered, enriched and routed centrally, the pipeline directly affects detection outcomes, compliance evidence and incident response. It must therefore be treated as production infrastructure, not a passive routing layer. The controls below focus on availability, confidentiality, governance and control-plane security to ensure telemetry remains usable, protected and traceable under normal operation and failure conditions.

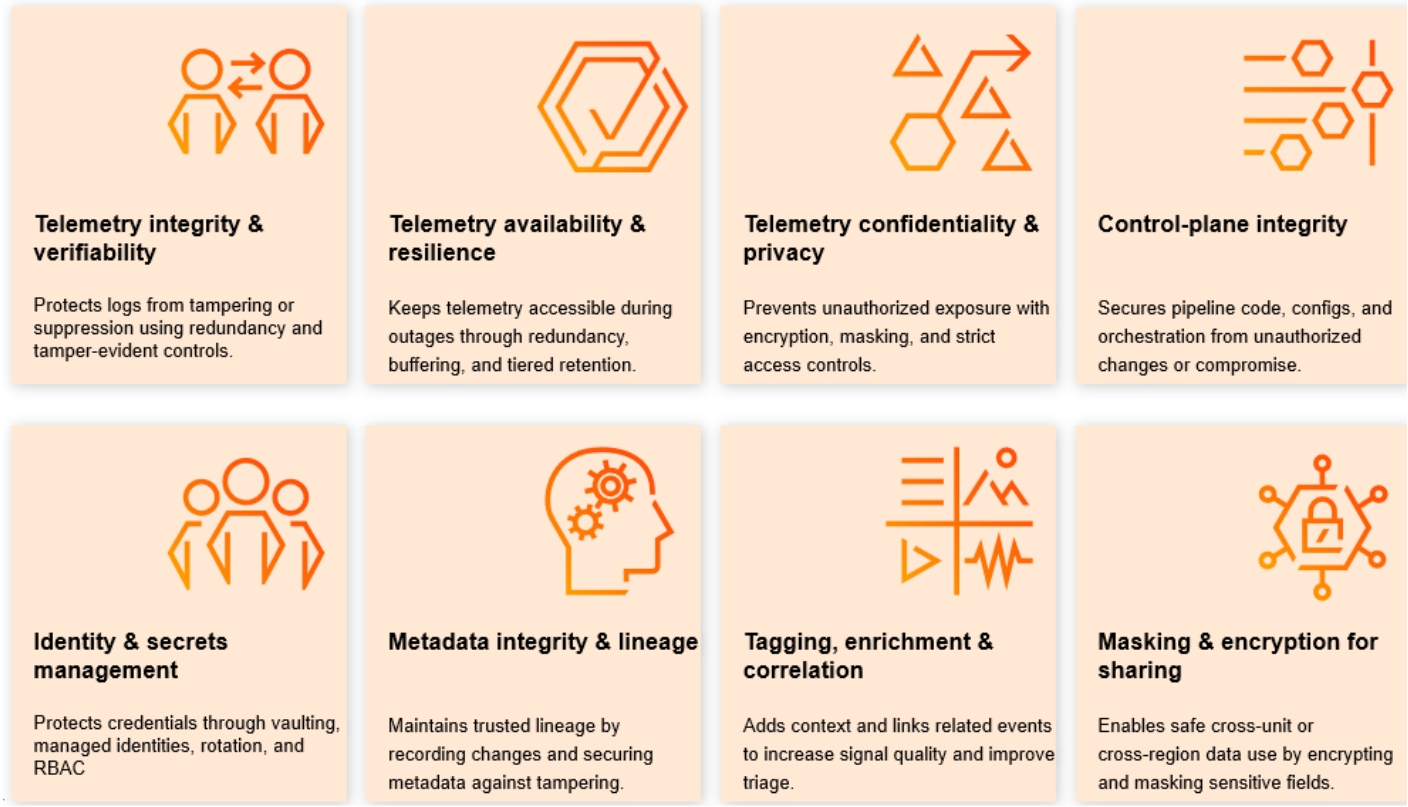
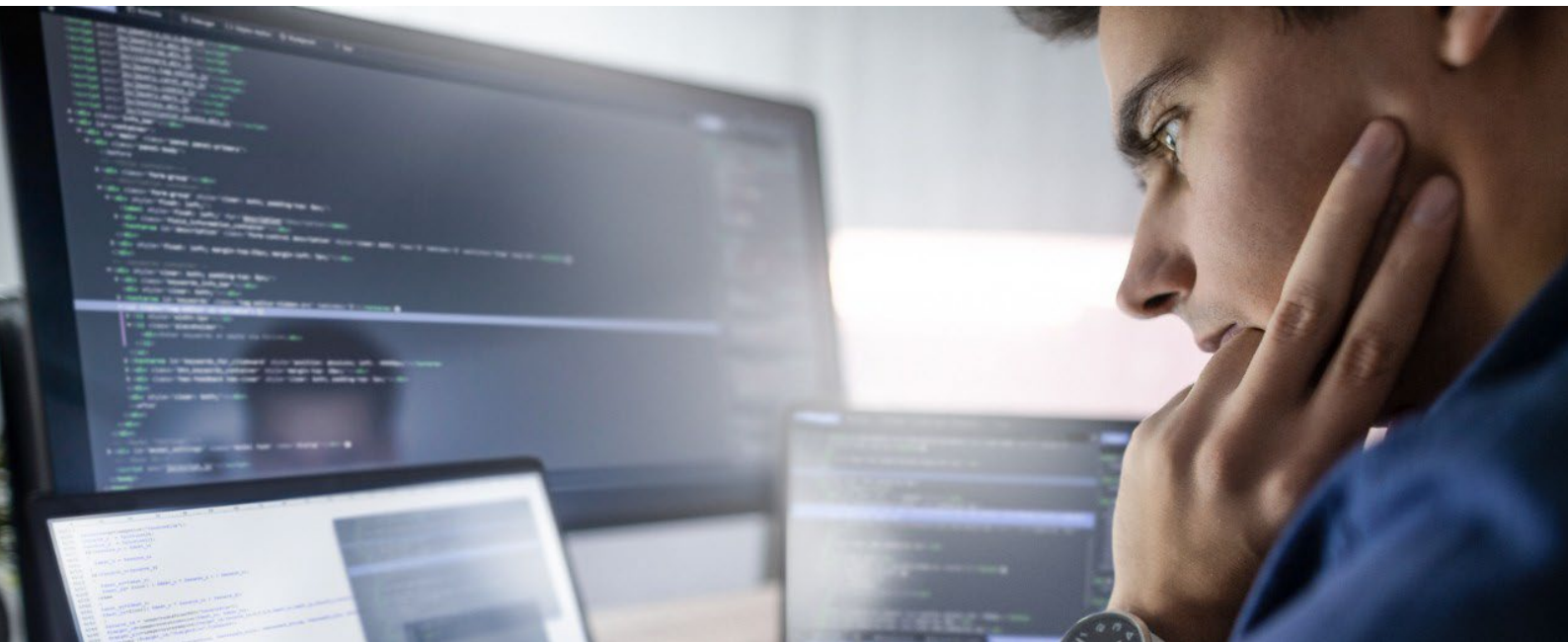


Figure 3: Overview of extended security use cases and control requirements

Future trends and considerations

- **Shift-left detection and agentic AI:** Security data pipelines are moving beyond transport to perform parsing, enrichment, and first-pass detection directly in the data stream [1]. Pushing detection logic upstream reduces latency, limits attacker dwell time, and prepares the ground for agentic AI that can autonomously tag, prioritize, or suppress low-value signals before they reach centralized platforms.
- **SIEM as an insight layer:** SIEMs are increasingly used as investigation and analytics consoles rather than as primary data stores [4]. Pipelines enable this shift by decoupling storage from detection and feeding SIEMs curated, high-value events instead of raw telemetry, improving detection efficiency while controlling cost [4].
- **Federated data mesh:** Rather than centralizing all telemetry, organizations are storing data where it is most cost-effective and querying it through federated search, especially in hybrid and multi-cloud environments [4]. Pipelines provide the consistency and routing logic needed to make this model operational without duplicating data.
- **Standardization and interoperability:** Adoption of open schemas such as OCSF is increasing to reduce fragmentation across tools and analytics platforms [4]. Aligning pipeline output to a common schema simplifies detection engineering, improves portability, and lowers long-term integration effort.
- **Shifting detection away from monolithic SIEMs:** Ingestion-based SIEM pricing is increasingly viewed as unsustainable at scale. Analyze-in-place and federated detection models avoid central data duplication by running detections where data already resides, reducing ingestion, storage, and rehydration overhead without reducing coverage.
- **Privacy and sovereignty:** As regulatory requirements tighten, pipelines are becoming enforcement points for data residency, sovereignty, and privacy controls. Embedding localization, encryption, and masking decisions into the pipeline allows global analysis while respecting jurisdiction-specific obligations.



Conclusion

Data pipeline management is not merely a technical undertaking but a strategic imperative. The explosion of security telemetry and the unsustainable economics of legacy SIEMs require a new approach. By adopting a purpose-driven security data pipeline, organizations can control costs, improve detection quality, support compliance and prepare for the AI-enabled future. The blueprint outlined here provides a foundation for building a resilient, scalable and business-aligned data architecture. Organizations that act now will be better positioned to defend against evolving threats, satisfy regulators and unlock the full value of their security data..

Sources: The insights and data in this paper are drawn from a range of industry studies, surveys, and expert analysis, including recent reports on SOC challenges, the integration of AI in security operations, vendor-neutral case studies on AI-native SOC outcomes, and forecasts by thought leaders on generative and agentic AI in cybersecurity. These sources have been cited throughout the document to provide supporting evidence for the discussed trends and recommendations..

Vishal Sharma

Director, Cyber Defense & Managed Security Services, PwC Germany

+49 1517 2931922

vishal.s.sharma@pwc.com

Himanshu Chaudhary

Director, Cyber Defense & Managed Security Services, PwC Germany

+49 1517 3059047

himanshu.chaudhary@pwc.com

This content is for general information purposes only and should not be used as a substitute for consultation with professional advisors.

© 2026 PricewaterhouseCoopers GmbH Wirtschaftsprüfungsgesellschaft. All rights reserved. “PwC” refers to PricewaterhouseCoopers GmbH Wirtschaftsprüfungsgesellschaft, which is a member firm of PricewaterhouseCoopers International Limited (PwCIL). Each member firm is a separate legal entity. Please see www.pwc.com/structure for further details.

Sources

1. SoftwareAnalyst, “Market Guide 2025: The Rise of Security Data Pipelines & How SIEMs Must Evolve,” SoftwareAnalyst Substack, Apr. 21, 2025.
2. Cybersecurity Ventures, Cybersecurity Almanac 2025 , Dec 11. 2025. (https://cybersecurityventures.com/cybersecurity-almanac-2025/)
3. A. Ganesan, “Security Data Pipeline Platforms: The Data Layer Optimizing SIEMs and SOCs,” DataBahn Blog, May 21, 2025.
4. HOOP Cyber, “Security Operations Predictions for 2026: The Year Security Data Architecture Reaches a Tipping Point,” HOOP Cyber Blog, Dec. 19, 2025.
5. C. DeRodeff, “Shift-Left Detections with Abstract,” Abstract Security Blog, Jul. 22, 2025.
6. GitLab, “Shift Left Security: A Complete Guide,” GitLab, 2023.
7. N. Tischler, “The Benefits of Shifting Left: Minimize Risk and Save Money with Early Security Integration,” Veracode Blog, May 13, 2025.
8. Confluent, “What Is Shift Left in Data Integration? Key Concepts & Benefits,” Confluent Blog, 2024.
9. Kusari, “Telemetry: What is it? Definition, Explanation & Data Collection Insights,” Kusari Blog, 2023.
10. Splunk, “What’s the CIA Triad? Confidentiality, Integrity, Availability,” Splunk Blog, 2023.
11. DataBahn, “Hybrid Data Pipeline Security: Best Practices for Telemetry in 2025,” DataBahn Blog, 2025.
12. Wiz, “CI/CD Pipeline Security Best Practices,” Wiz Research, 2024.
13. Microsoft, “Best Practices for Protecting Secrets,” Microsoft Learn, 2023.
14. BIX Tech, “Audit-Friendly Data Pipelines: Capture Lineage, Metadata & Governance,” BIX Tech Blog, 2024.
15. Conifers AI, “Security Telemetry Data Pipeline (Glossary),” Conifers AI, 2024.
16. Enterprise SIEMs Miss 79% of MITRE ATT&CK Techniques Used by Adversaries, According to CardinalOps' 5th Annual Report , June 2025