White Paper

# Federated Data Management and Search: The Next Frontier of Data Management for Digital Resilience, Optimization, and Competitive Advantage

Sponsored by: Splunk

Archana Venkatraman
June 2024

## SITUATION OVERVIEW

Technology and business leaders are primarily concerned about cybersecurity, impactful AI investment, and ROI.

Cybersecurity is at the forefront of C-suites' minds because it is getting increasingly complex forcing organizations to progress from reactionary approaches to proactive, insights-driven strategies. According to IDC's December 2023 *Future Enterprise Resiliency and Spending Survey,* about 65% of organizations said they experienced a ransomware attack in 12 months and 40% of affected organizations admitted to paying a ransom to regain access to their own systems and data.

It is no surprise that the top-most IT function immune to budget reduction in 2024 is security, risk, and compliance (source: IDC's *Future Enterprise Resiliency and Spending Survey, Wave 2,* February 2024).

Additional research from IDC in 2023 indicated that 6 in 10 organizations expect to increase their spending on cyber-recovery tools and solutions over 24 months.

But the threat landscape is changing rapidly driven by multiple factors such as sophisticated attack patterns driven by AI, a broadening threat surface area, and growing IT complexity and silos. To add to this, 8 in 10 organizations operate in hybrid cloud environments with ever more distributed environments to manage. In IDC's February 2024 CISO Hub discussions, one CISO reported their organization had already suffered a deepfake attack. Another mentioned past incidents involving AI-generated voice cloning in spear-phishing attacks.

Modern cyberthreats are fundamentally challenging traditional security strategies. When the stakes are so high, status quo strategies are falling short. This is because traditional strategies have introduced silos of data. Security teams, IT operational teams, and recovery teams all operate in silos, leading to gaps in insights.

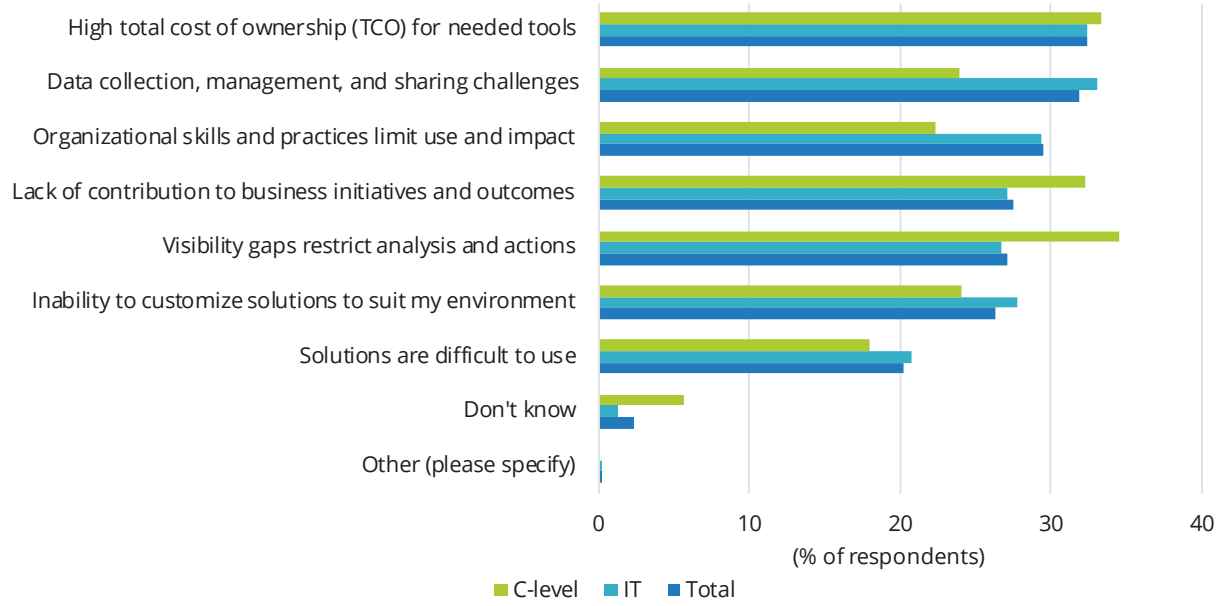
Tool sprawl is leading to data sprawl that results in visibility gaps. According to IDC's November 2023 *Future Enterprise Resiliency and Spending Survey, Wave 10,* 37% of organizations admitted to having 6–10 discrete observability solutions (including on-premises systems, cloud-based SaaS, and managed services) and another 25% of organizations said that they have more than 10 discrete observability solutions.

*The biggest barriers to gaining maximum benefit from observability strategies are visibility gaps, high costs, and limited effectiveness (either through limited contribution to business needs or the lack of flexibility or complexity).*

This is highlighted in Figure 1.

## FIGURE 1

### Biggest Barriers to Success in Gaining Maximum Benefit from Observability Solutions




n = 881

Source: IDC's *Future Enterprise Resiliency and Spending Survey, Wave 10,* November 2023

Interestingly, C-level leaders are most worried about the visibility gaps (refer back to Figure 1) and limited business impact from investments. IT leaders are facing obstacles around data collection, management, and sharing, while high costs remains the number 1 challenge overall.

Siloed strategies result in a reactive and delayed approach to risk management, and more importantly, it puts organizations at a competitive disadvantage because they are not able to effectively leverage insights to have an evidence-based approach to resilience. In conversations with IDC, C-level executives highlighted challenges such as "dashboard fatigue," "disruption due to false alerts," and "resource wastage" coming from the lack of convergence between security and observability.

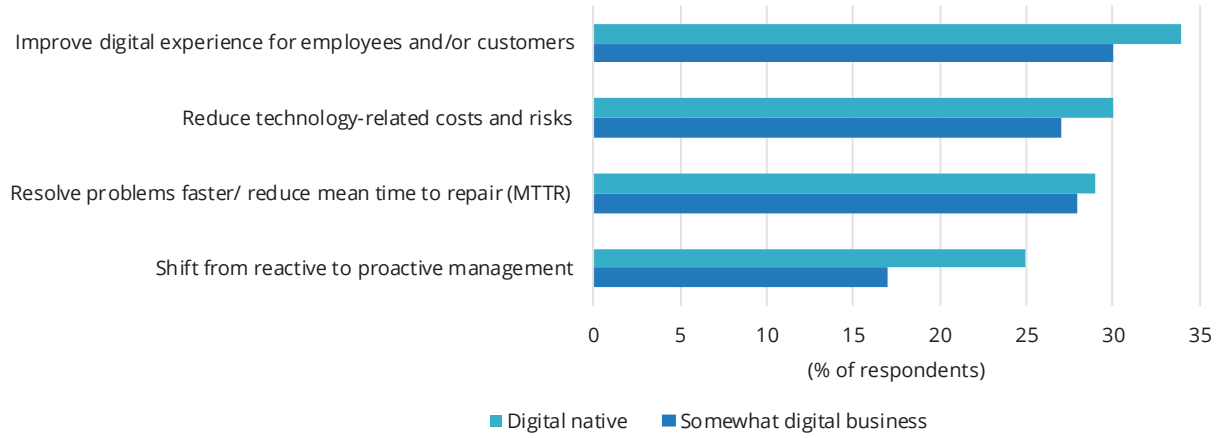
## Building the Case for Convergence and Taking a Data-Driven Approach to Resilience

IDC research shows that only a handful (18%) of global organizations have a highly observable technology environment and provide visibility into business outcomes that reach the C-suite. The number is even lower in regions such as EMEA and Asia/Pacific.

But the expectations of benefits from advanced observability solutions are high. IDC's data shows that those that apply advanced observability capabilities and take a data-driven approach are reaping meaningful benefits (see Figure 2). Digital natives overall are able to scale the benefits more, become proactive in digital resilience, and are able to reduce risks and costs more effectively than organizations that are less advanced.

## FIGURE 2

### Expected Business Outcomes from Advanced Observability Strategies and Solutions




n = 881

Source: IDC's *Future Enterprise Resiliency and Spending Survey, Wave 10,* November 2023

And success begets success. Because they see better value from investments, a higher percentage of digital natives (76%) agree or strongly agree that a comprehensive, unified solution drives productivity, efficiency, and innovation compared with 64% of organizations that are somewhat digital.

## *Honing the Data Strategy to Boost Business Resilience*

The foundation for strong cyber-resilience and better security outcomes is having a data-centric strategy. While most organizations collect data, without effecting management, it is difficult to turn it to insights. IDC research from 2020 has estimated that as much as 68% of data goes unleveraged primarily because of lack of data management. With the growing volume, velocity, and variety of data, this problem is further amplified, increasing the importance of data management. Augmenting with a data management strategy through a unified platform combining security and observability helps deliver a huge competitive advantage for resilience. In IDC's opinion, one of the biggest advantages of combining security and observability data in a single platform is that it helps organizations surface and analyze the data with richer context as they are able to better correlate and contextualize the findings. It can also help them overcome the key challenges around visibility islands and conflicting data signals.

Data collected through rich telemetry logs and metrics traces and their contexts collected across the hybrid digital footprint has tremendous value spanning across multiple use cases throughout its phases:

For instance:

- Real-time data is valuable in preventing, detecting, and monitoring attacks.
- Over time, as data is moved to archive, it serves compliance, audit, and forensic use cases.
- Beyond real time, data is valuable for additional ad hoc use cases such as "what-if" analyses, scenario planning, threat hunting, pattern insights, investigation, and incident review and learning.

## FUTURE OUTLOOK

While data has value across multiple stages of its life cycle, optimizing it and capitalizing on it are challenging without a unified data management strategy that focuses on the key principles discussed in the sections that follow.

### Smarter Management and Tiering of Data for Cost Optimization

Real-time data serves critical use cases such as detection, meeting critical SLAs, and improving mean time to resolution. This data needs to be highly available and quickly accessible. This requires it to reside on high-performing, low-latency environments.

Meanwhile, older data serves other audit and compliance or threat analysis use cases but need not reside in expensive tier 1 storage environments. Storing all data on primary, high-performant environments increases the cost of infrastructure; has management overhead; and increases complexities. It is likely to also affect the performance and responsiveness of real-time data because of the growing data density on the infrastructure.

Policy-based tiering of data into cost-effective storage can help address the challenges mentioned previously. However, it is essential to make sure that data in cost-effective environments remains accessible when the need arises. Without the ability to access that data in realistic timescales and costs, archiving strategy cannot serve the secondary use cases.

A unified approach to data management can help organizations store the right data in the most efficient infrastructure, scale cost effectively, avoid replication of data, and improve the speed of insights to deliver impactful business outcomes. The modern strategy is to develop a unified "approach" to data management and not necessarily unify all the data. IDC's research since hybrid cloud and multicloud became a norm indicates data mobility is one of the most challenging, inefficient, and cost-intensive projects. Further:

> The value of a unified approach to data management is the ability to maintain federated access to data as it ages and across data stores to benefit from an appropriate price-to-performance ratio.

With this federation-defined approach, the emphasis is on unifying access (where value resides) and not on unifying data repositories (which is time, skills, and cost intensive).

#### *Data Pipeline Management*

With hybrid cloud and system diversity becoming a norm, operational data is high in volume, variety, velocity, and veracity. Without the ability to transform and standardize this data when needed, the quality of insights can be poor, manual, and resource intensive. A data pipeline architecture provides

the ability to clean, schematize, filter, and transform data during ingestion, before sending it to its final destination ready for use. Efficiency, speed, and data quality from modern data pipelines determine the success of real-time data analysis.

Increasing automation and orchestration across data pipelines, applications, and infrastructure was cited by 46% of organizations as the most important step to ensure IT systems can fully support data-driven business decisions (source: IDC's *Future of Digital Infrastructure Survey,* 2023). The ability to improve, automate, and modernize data pipelines is high on the priorities for organizations as they develop mature IT practices. IDC believes that automation and orchestration is a future state, and the primary focus must be first on modernizing the data management approach as a prerequisite for automation.

## *Federation of Data*

A unified approach to data management for visibility is one half of an effective data-centric strategy. It helps create different layers of data ready for use cases. The second half of the data-centric strategy is federation of data, especially around its access and in-context analysis. This aspect brings in flexibility and choice for IT leaders to store data where they want — on premises, SaaS, or public cloud infrastructure data lakes.

Moving data to a centralized location to serve every use case is a costly, time-consuming, and inefficient strategy.

A centralized data lake approach can bring risks such as single point of failure, high costs, and limited scale.

According to IDC's 2024 *CloudOps and Cloud Governance Survey,* data environments are cited among the ones that are most underutilized or "wasted."

Among the main underutilized resources are:

- Cloud security and cloud databases (cited by 31% of respondents)
- Cloud data warehouses (cited by 26% of respondents)
- Cloud data lakes (cited by 22% of respondents)

This is where data lake federation becomes a valuable strategy because data stays in its original location. Exposing data in security data lakes using scanners or APIs can provide visibility without moving or re-ingesting data each time. This federated visibility across existing, familiar environments improves the speed of analysis and unlocks extemporaneous and long-term security use cases such as forensic investigation or comparing data for auditing use cases. With traditional approaches, organizations admit that their forensic investigations were just tabletop exercises because iterating them for meaningful analysis was becoming an expensive process.

The urgency of resolving this is translating into targeted investments, especially by savvy organizations. Nearly 40% of digital-native organizations said they will "significantly increase engagement" with external providers specifically for areas of cyberevent responses such as risk analysis, cyberposture, incident response, forensic analysis, and recovery orchestration. In comparison, only 26% of non-digital natives said the same (source: IDC's *Future Enterprise Resiliency and Spending Survey, Wave 10,* November 2023). Data federation allows organizations to execute on

forensic analysis without overrunning budgets and not be afraid of using data for multiple security use cases.

The ability to unify data management and federate visibility and access can be a game changer, especially as system diversity and data increase with more IoT, edge, and mobile devices on the rise within organizations. In IDC's opinion, large organizations, especially in heavily regulated environments, can unlock several use cases and boost their competitive advantage with a data federation strategy. Key benefits include:

- Minimized risks around privacy, security, or data misuse
- Real-time learnings
- Accelerated model training or AI strategies with data
- Accelerated innovation with broader and diverse systems

Beyond that, savvy organizations need to consider data federation as an imperative because the relevance of data federation is growing amid rising sovereignty sentiments. Furthermore:

> IDC predicts that by 2024, 50% of G2000 organizations will need to demonstrate compliant digital sovereignty practices aligned with local legislation on data processing, storage, classification, and management.

High sovereign sentiments mean that organizations will become more reluctant to move data to a central location and instead prefer scanners and data-exposing mechanisms to leverage the data in secure repositories in their geolocations. According to IDC's January 2023 *Future Enterprise Resiliency and Spending Survey, Wave 1,* 60% of organizations across the world said they were either already currently using or considering using sovereign cloud solutions in the next two years.

## Discoverability of Relevant Data (Impactful Search Capabilities)

The final piece of a modern data management strategy is intelligent search. The business value of search lies in how it empowers teams to find relevant data quickly and reuse it for different use cases such as other AI/ML projects, model training, scenario testing, recovery testing, compliance and audit, and analytics. In fact, 38% of cloud leaders and 30% of LOB leaders cited "cross-platform data search and visibility" as their most desired data capability in IDC's 2024 *CloudOps and Cloud Data Management Survey.* Data findability was more important than other key data capabilities such as data integration or data APIs. Reusing data not only unlocks multiple new use cases but also helps businesses sweat their data assets for business value. For a long time, organizations have found it difficult to balance data search and discoverability with cost efficiency. Savvy organizations are seeking solutions that help them search for data at the right cost-to-performance ratio (i.e., have a data platform that can support granular search in natural language and surface relevant, high-quality data from any environment without a knock-on cost effect). Successful data federation includes insightful search and analytics on that federated data.

Together, unification of security and observability data, modern data pipeline management, and federated search and analytics holistically address the three most important C-suite needs:

- Cyber-resilience with insights-driven decisions
- Improved ROI or impactful value from investments
- Readiness of data for accelerated AI innovation

At a business level, combining security and observability data fortifies resilience because insights from observability platforms strengthens security practices and improves quality of developer services. At an operational level, a data pipeline management platform infused with federated search surfaces mission-critical data in seconds. And finally, from a culture perspective, it improves collaboration between IT, security, data, and developer teams and gives them a common lingua franca for success.

## RECOMMENDATIONS AND CONCLUSION

The threat of sophisticated breaches and the global spotlight on data privacy and resilience require IT leaders to strengthen their organizations with robust, agile, and preventative cybersecurity measures.

Security compliance is an area of no compromise for enterprises, but it is dauntingly difficult, rigid, and expensive. Organizations need a modern data management strategy that helps plug visibility gaps in resilience and improve security postures without costs spiraling out of control or additional complexity.

Flexibility, data optimization, and data reuse must be embedded in data management platforms.

As organizations seek to modernize their security, observability, and data management strategies, they need to take a holistic approach and evaluate a platform that helps converge security and observability data and supports multiple use cases through data pipeline and federation capabilities. Exploring the value of a platform that combines expertise in SIEM, SOAR, observability, and data management capabilities can provide organizations with a golden path to success and to develop a competitive advantage.

### Evolving from Product to Platform

IDC expects that in the next 20–24 months, at least 20% of organizations will use a proactive cybersecurity platform that aggregates risk exposures to score and prioritize cybersecurity risk in totality instead of one tool at a time.

Multiple point products and the data and signals coming from them are harder to correlate and manage. A platform approach can help in presenting clearly, transparently, and accurately, so the time to action is accelerated. IDC believes that a holistic platform-based view delivers a better and faster understanding of scenarios for quick remedial action.

Organizations need to choose a unified platform that is open, integrated with hybrid cloud architectures, and extensible. They also need to make sure the platform they choose can scale and has a track record of innovation in the fields of security, observability, and data management as well as a vision to leverage new AI innovation to improve its services. For example, search and analytics federation brings huge opportunities for organizations today. But in the long run, the platform should evolve and infuse advanced AI capabilities to add richer context to a search query, boost privacy, automate redaction of sensitive data, and personalize search results for different personas based on the search context (developers, ITOps, SOC teams, cloud engineers, C-suites, risk and legal teams, etc.)

Data is the currency of a modern digital business, and an intelligent data management strategy holds the key to proactive security and digital resilience amid unprecedented threats.

*Are you ready to optimize the value from security data to unlock new levels of digital resilience?*

## About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

## Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com

---