zscaler™

WHITE PAPER

Data Fabric For Security — What it Is, and Why it Uniquely Addresses the Security Data Challenge

Contents

Introduction 3

Part 1 – Understanding the Data Fabric 4

What is a data fabric? 4

Why is a data fabric needed to do security and risk management well? 5

How does data fabric differ from a data warehouse or data lake? 6

How does a data fabric differ from other security architectures like SIEMs? 9

What factors should you consider when thinking about building a security data fabric? 11

Part 2 – The Zscaler Data Fabric for Security 13

How did the founders approach building a data fabric differently? 13

How does the Zscaler Data Fabric for Security work? 14

What makes the data model at the heart of the Zscaler Data Fabric for Security unique? 15

What solutions does the Zscaler Data Fabric for Security power? 16

Conclusion 17

Glossary of Terms 19

Introduction

In an era defined by digital transformation, data has emerged as both an asset and a challenge for organizations. The rapid adoption of cloud computing, IoT, and AI technologies has led to explosive growth in data volume, variety, and velocity. This shift has created complex, multicloud, and hybrid environments, making it increasingly difficult for enterprises to integrate, manage, and leverage data effectively. Traditional data management solutions, such as data warehouses and lakes, were designed to store and analyze data but they fall short of providing seamless integration, real-time analytics, comprehensive governance, and intelligence.

To deliver on these more sophisticated requirements requires a data fabric. A data fabric is an architectural framework that unifies disparate data sources, enabling a consistent, real-time layer of data management across environments. By simplifying access, enhancing governance, and supporting rapid analytics, it provides organizations with a strategic edge, enabling faster and more informed decision-making. In fact, in a Top Strategic Technology Trends for 2022 report Gartner estimated that data fabric deployments will reduce data management efforts by up to 70% and accelerate time to value.

In the realm of cybersecurity, where real-time intelligence and rapid threat detection are crucial, data integration challenges are even more pronounced. Security teams often face a deluge of unintegrated data from various tools, leading to fragmented visibility and slower response times. Zscaler's innovative approach to data fabric architecture tackles this issue head-on, offering a security-focused solution that normalizes, enriches, and correlates data for better insights and faster action.

This whitepaper explores the fundamentals of data fabrics, their role in security, and how Zscaler has built its Data Fabric for Security to redefine security operations. From driving real-time decision-making to improving risk prioritization, the Zscaler Data Fabric for Security exemplifies the transformative potential of data fabrics in modern cybersecurity strategy.

PART 1

Understanding the Data Fabric

What is a data fabric?

A data fabric is a comprehensive solution for managing data in today's multicloud, hybrid, and complex data ecosystems. It simplifies data access, enhances governance, and supports real-time analytics, making it an essential framework for organizations aiming to leverage data as a strategic asset. Data fabrics streamline operations, improve decision-making, and ultimately drive better outcomes.

But what exactly is a data fabric? In short, a data fabric is an architectural framework that enables the integration, management, and delivery of data across diverse and complex environments. It weaves together disparate data sources, whether on-premises, in the cloud, or in hybrid environments, providing a unified and consistent layer of data access and governance. This approach addresses the growing complexity of managing data as organizations deal with vast amounts of structured, semi-structured, and unstructured data from a wide array of sources.

While data fabrics are utilized to drive many information technology outcomes, they have a common set of underlying capabilities:

Unified data access: Data fabric integrates data across multiple platforms and formats, creating a single environment where data can be accessed regardless of where it resides. It abstracts the underlying infrastructure, allowing users and applications to interact with data without needing to know where it is physically stored.

A metadata-driven approach: Data fabric relies heavily on metadata (data about data) to automate data management tasks. This metadata-driven approach facilitates intelligent data discovery, governance, and access. By understanding relationships between data assets, the fabric can enable dynamic data orchestration, ensuring that the right data is available at the right time.

Data governance and security: A data fabric enforces consistent governance and security policies across the entire data landscape. This ensures that data privacy and regulatory requirements are met, regardless of the source or location of the data. The framework allows for centralized policy management, but it is also flexible enough to apply localized controls where necessary.

Real-time data access and analytics: Data fabrics are designed to support real-time or near-real-time data processing and analytics. This is crucial for modern businesses that rely on up-to-date insights for decision-making. The ability to access and analyze data in real-time allows organizations to respond quickly to changing market conditions, customer behaviors, or operational challenges.

Why is a data fabric needed to do security and risk management well?

In today's world, cybersecurity is synonymous with data. Security solutions aim to protect both personal and business data, generating a vast amount of data in the process. It's no secret that security leaders grapple with a flood of disjointed data streams, lacking the resources or expertise to integrate and interpret it effectively.

It was not so long ago that security teams longed for more data. Early tools provided basic security metrics and enterprise security posture reports. However, with increasing demands for risk intelligence and real-time reporting, CISOs sought solutions delivering more data-driven insights. They wanted assurance that their tools were effective.

The market responded, perhaps too enthusiastically. Over the past two decades, a multitude of security solutions—from MDRs to vulnerability scanners and threat intelligence platforms—have emerged. Large enterprises today often deploy over 30 different tools in their security stack, each generating its own data. However, this data, unique to each tool's purpose, often doesn't integrate or de-duplicate with data from other solutions, leading to a paralyzing amount of disconnected information. Security experts, not typically data experts, face an uphill battle to integrate these data streams and gauge their security function's performance.

Unfortunately, these tools and programs have not lived up to their promise because security teams have struggled to extract value and context from their data. Furthermore, the data they do have lives in silos, hindering the ability to gain a unified and comprehensive view of security threats and vulnerabilities.

By implementing data fabric architecture, security teams can integrate diverse data sources, creating an open, scalable foundation for numerous use cases. This flexible system accommodates multiple data points from across the enterprise, offering reliable and actionable insights.

A source of truth: One of the main reasons cybersecurity teams need a data fabric is to eliminate the fragmentation caused by the wide array of security tools and vendors. Unlike other industries where a single system (like Salesforce in sales) can provide a holistic view, cybersecurity operates with numerous specialized tools, each focusing on different aspects of security—such as vulnerability management, asset management, and security information and event management (SIEM). These tools often store their data in separate silos, making it difficult to gain a complete and accurate understanding of the security landscape.

A data fabric serves as a unifying layer that integrates data from all these tools, creating a "single source of truth." It translates and correlates data from various systems, making it easier for teams to search, analyze, and derive insights across their entire security environment. By providing a consolidated view, a data fabric enables cybersecurity professionals to understand the relationships between different data sources and entities, such as users, assets, and vulnerabilities, leading to more effective threat detection and response.

Enhancing real-time decision-making: The speed at which cyberthreats evolve necessitates real-time decision-making. However, without a unified data architecture, teams often spend considerable time piecing together data from different sources, which can delay response times and allow threats to proliferate. A data fabric accelerates decision-making by normalizing and deduplicating data in real-time. It creates a unified, enriched view of the security data, allowing teams to quickly understand what is happening across their network without needing to manually correlate information from multiple systems.

Moreover, a data fabric supports automated enrichment and entity resolution. For example, it can automatically resolve data points from different sources, ensuring that information about a particular asset or vulnerability is aggregated and cross-referenced correctly. This reduces the likelihood of errors, such as missing a critical vulnerability due to incomplete data. By resolving and correlating data from different systems, a data fabric provides cybersecurity teams with a clearer and more accurate understanding of their security posture, allowing them to make faster, more informed decisions.

Improving risk prioritization accuracy: In cybersecurity, accurate and timely detection of threats is crucial. However, traditional SIEM systems, which primarily aggregate raw data for threat detection, do not provide the contextualized insights necessary for comprehensive security analysis. A data fabric enhances threat detection by transforming raw data into actionable intelligence. It enriches the data by correlating it with external sources, such as threat intelligence feeds, and internal sources, like user and asset data, to provide a more comprehensive view of potential threats.

Additionally, a data fabric enables more effective cross-correlation of data, allowing cybersecurity teams to detect patterns and relationships that would otherwise be missed.

For instance, it can link a user's activity data with known vulnerabilities and external threat intelligence, providing a more nuanced understanding of potential risks. This level of correlation is crucial for identifying complex, multi-vector attacks that may not be detectable by analyzing individual data points in isolation.

How does data fabric differ from a data warehouse or data lake?

A data fabric, data lake, and data warehouse each play unique roles in managing data within an organization. While they all serve as part of the broader data architecture, their designs, purposes, and the types of data they handle differ significantly. Here's a closer look at how these three solutions compare.

Data fabric: Unified and real-time integration across sources

A data fabric is a holistic data architecture designed to unify and integrate data from multiple sources, whether on-premise, in the cloud, or at the edge. It enables a consistent, real-time view of data across an organization by providing a unified access layer. Unlike data lakes or data warehouses, a data fabric does not focus solely on data storage; instead, it emphasizes seamless data integration, governance, discovery, and orchestration. This approach reduces data silos and allows for better accessibility and decision-making across departments.

The data fabric integrates structured, semi-structured, and unstructured data, making it accessible in real-time for a broader range of users—not just data specialists.

Much of this data is often stored in a data warehouse or data lake, but data fabrics take this critical data and apply advanced analytics, machine learning, and automation features to enable rapid processing and contextualization of data from these diverse sources. The architecture is designed to support complex processes, enhance data security, and ensure compliance by enforcing access controls and governance policies. However, implementing a data fabric on your own is complex, requiring specialized skills, a well-planned integration strategy, and significant investment. Hint: the good news is that Zscaler has done all of this work for you. You just get to reap the benefits!

Data warehouse: Structured data for predefined analysis

A data warehouse is designed to store structured and refined data that has been cleaned, transformed, and organized into predefined schemas. It supports high-speed analytical processing and is commonly used for business intelligence, reporting, and trend analysis. Data warehouses consolidate data from various sources using the Extract, Transform, Load (ETL) process to ensure consistency and accuracy.

Data warehouses excel in delivering rapid, reliable insights from processed data, making them vital for operational decision-making. They are well-suited for running complex SQL queries and generating reports for management and business teams. However, they lack the flexibility to store unstructured data or handle experimentation with raw data, limiting their usefulness for emerging analytics or exploratory processes.

Data lake: Centralized raw data storage

A data lake is a centralized repository that stores massive amounts of raw data, including structured, semi-structured, and unstructured data, in its original format. It is designed to handle vast amounts of diverse data, ranging from log files and IoT sensor data to audio, video, and social media feeds. The primary function of a data lake is to provide flexible and cost-effective storage until the data is ready for processing and analysis.

Unlike a data warehouse, a data lake is not preconfigured with a specific schema, making it more suitable for experimentation and advanced analytics like machine learning and AI. Data lakes often require technologies like Hadoop, Apache Spark, or NoSQL databases to manage and process the diverse data types they contain. However, due to its raw nature, a data lake can quickly become a "data swamp" if not properly managed with metadata, quality control, and governance measures. Users also need specialized skills to extract value from data lakes, as querying raw data can be complex and time-consuming.

The data quality hierarchy

When looking at different data architectures, it's important to consider the quality of the data within, and how the data will be used. One good way to look at this is often referred to as the data quality hierarchy. The "Gold," "Silver," and "Bronze" data layers represent a hierarchy of data quality, refinement, and readiness for use within a data architecture. These layers are essential components in managing and optimizing data workflows, ensuring that organizations can efficiently extract value from their data.

The Gold layer consists of the most refined, high-quality, and analytics-ready data. Data in this layer is clean, transformed, and fully curated, with standardized formats and metadata. It is typically derived from the Silver layer and is used for critical decision-making, advanced analytics, and business intelligence. The data here is often aggregated or summarized to ensure it meets the highest standards for accuracy, consistency, and usability. It is suitable for dashboards, reports, and machine learning models, where precision is crucial.

The Silver layer represents an intermediate stage of data processing. It contains partially refined data that has undergone initial cleaning, transformation, and validation. While not fully polished, Silver data is more usable than raw data, making it ideal for exploratory analysis and preliminary reporting. It often includes data that retains some level of granularity, enabling deeper investigation and troubleshooting before being elevated to the Gold layer. It acts as a bridge, transforming raw data into more usable forms without complete aggregation.

The Bronze layer consists of raw, unprocessed data ingested from various sources, often in real-time or batch loads. This data is stored in its original form, making it the least refined. It serves as a foundational layer for all other processing and provides a complete record of collected data. While Bronze data is not immediately usable for analytics, it's essential for historical analysis, anomaly detection, and debugging.

These layered structures ensure flexibility, scalability, and better governance in data management.

	Data Fabric	Data Warehouse	Data Lake
Purpose	To provide data integration, governance, and accessibility for all users, creating a single source of truth and making data less siloed for multiple use cases.	To support high-speed analysis of historical data for reporting and business intelligence.	To provide a large-scale, cost-effective repository for raw data, ideal for exploration and advanced analytics.
Data types supported	Structured, semi-structured, and unstructured data across all sources.	Only structured, processed data with predefined schemas.	Raw data in all formats without predefined schemas.
Data quality hierarchy	Gold	Silver	Bronze

How does a data fabric differ from other security architectures like SIEMs?

A security data fabric and a Security Information and Event Management (SIEM) serve different roles within the cybersecurity ecosystem, with each having a distinct approach to handling data. The main difference lies in their core objectives and architecture.

Data storage vs. data optimization

One of the most significant distinctions between a SIEM and a security data fabric is their approach to data storage. Traditional SIEMs are built around the idea of aggregating and storing large amounts of log data. This massive data storage drives the SIEM's revenue, as it is often tied to the volume of data processed and stored. The more data collected, the higher the consumption and licensing costs, creating an incentive to store as much data as possible.

In contrast, a data fabric is designed to optimize and manage data efficiently. It focuses on reducing data volume by normalizing, enriching, and deduplicating information before routing it to various destinations. A data fabric enables organizations to send data to lower-cost storage options, further reducing storage expenses. This optimization not only improves data quality but also ensures more flexibility in handling different types of security data across the enterprise. The key challenge here is that SIEM vendors have little incentive to prioritize data volume reduction, as it contradicts their revenue model.

Events vs. entities

SIEMs are fundamentally event log-driven. They collect, aggregate, and analyze log data in near real-time to identify anomalies and security threats. Each event log represents a discrete action or incident, such as a login attempt, file access, or system error. The primary purpose of a SIEM is to detect potential threats by examining patterns in these logs. As such, SIEMs excel in scenarios where rapid log analysis and event correlation are required. Their focus remains on capturing individual events and assembling them to generate alerts, making SIEMs particularly effective for real-time monitoring, incident response, and forensic investigations.

Data fabrics, on the other hand, are entity driven. Instead of treating data as isolated events, a data fabric seeks to connect related pieces of information to form a cohesive understanding of entities such as devices, users, applications, or assets. This approach enables a more holistic view, where data is not only collected but also enriched, normalized, and correlated around core entities. For example, a data fabric will merge data from multiple sources to provide a single, unified view of an asset, user, or vulnerability. By focusing on entities and their relationships, data fabrics offer deeper insights into security risks, asset management, and compliance.

A SIEM's event log-centric approach makes it ideal for detecting and responding to specific incidents, while a data fabric's entity-driven perspective provides broader context, supporting long-term security and operational needs. This differentiation makes data fabrics a vital part of a more adaptive and comprehensive security strategy.

Proprietary formats vs. interoperability

Another major difference is in data formats and integration. SIEMs are typically built on proprietary data formats, which can make it challenging to support interoperability. As a result, SIEMs may struggle to integrate smoothly with other tools or manage data formats from competing SIEM products. While there have been efforts to standardize data formats (e.g., through the Open Cybersecurity Schema Framework), achieving full compatibility remains a long-term goal.

Security data fabrics, on the other hand, are explicitly designed to be interoperable, working with various data formats and destinations. This capability is crucial for organizations implementing a multi-SIEM or security data lake strategy, where flexibility and support for multiple data formats are essential. The fabric's ability to unify disparate data sources makes it more suited to address complex security challenges, providing a holistic view that SIEMs cannot achieve.

Vendor lock-in vs. connecting to everything

A primary motivation for using a security data fabric is to avoid the vendor lock-in commonly associated with SIEMs. SIEMs, due to their complex integrations, often become deeply embedded in an organization's security operations, making it costly and time-consuming to switch vendors. SIEM products are unlikely to promote easy migration or integration with competitors, as doing so would undermine their market hold.

In contrast, a data fabric is built to be security data agnostic, enabling seamless data movement across various tools and solutions. This flexibility empowers organizations to switch solutions more easily, leveraging the best technologies available without being confined to a particular SIEM's ecosystem. By decoupling data collection and ingestion from the core functions of threat analysis, detection, and forensics, a data fabric maintains flexibility while strengthening overall security architecture.

While it's reasonable to question the convergence between SIEMs and data fabrics, it's crucial to understand that they can be complementary rather than competing technologies. SIEMs are excellent at real-time monitoring, alerting, and forensic analysis—core functions that they are optimized to perform. Data fabrics, however, excel at data preparation, integration, and optimization, serving as the "single source of truth" for all security data.

	Data Fabric	Data Warehouse
Purpose	To enable data integration, governance, and accessibility for all users, making data less siloed and enabling real-time insights.	To capture individual security events and assemble them to generate alerts.
Data types supported	Structured, semi-structured, and unstructured data across all sources.	Event logs for real-time monitoring, alerting, and query analysis.
Data quality hierarchy	Gold	Silver

What factors should you consider if you're thinking about building your own security data fabric?

Building your own security data fabric might seem appealing to organizations seeking complete control over their data infrastructure. However, the complexity, cost, and ongoing demands of such an endeavor make it a risky and potentially flawed strategy. Here's why attempting to create a custom-built security data fabric can pose significant challenges.

Do you have the necessary specialized expertise in house?
Developing a security data fabric is not like creating a typical data pipeline or implementing a data warehouse. A data fabric is a sophisticated, multi-layered architecture designed to integrate, normalize, and correlate vast amounts of diverse data from multiple sources, whether structured, semi-structured, or unstructured. Building such a system requires extensive expertise in data integration, governance, security, and machine learning. Organizations must also understand the intricate details of various data formats, security protocols, and APIs.

Achieving this level of integration and functionality is challenging. It involves custom engineering at every level, from data ingestion and entity resolution to enrichment and deduplication. In many cases, even organizations with advanced data science teams lack the broad and deep expertise necessary to build a fabric that can scale effectively across different use cases and environments.

Are you prepared for the hefty price tag to build and maintain?
Building a security data fabric from scratch demands significant upfront investment in both technology and personnel. Organizations need to allocate funds for infrastructure, advanced software, and hiring (or training) specialized staff like data engineers, architects, and cybersecurity experts.

Even after the initial build, the costs continue. Data sources and formats frequently change, requiring constant updates, maintenance, and optimization of connectors, parsers, and entity models. New security tools, regulations, and data standards also require continuous adjustments. Without the right resources, maintaining a custom-built fabric becomes increasingly expensive and time-consuming, potentially leading to technical debt and a loss of agility.

How will you manage scalability and flexibility?
Data fabric solutions must be designed to scale—both in terms of data volume and functionality. Off-the-shelf data fabrics are typically built to handle complex data transformations, high-volume ingestion, and rapid analytics across petabytes of data. Attempting to achieve this level of scalability internally can result in performance bottlenecks and inefficient processing. Custom-built solutions often struggle to maintain performance as data volume grows, leading to delays in insights and increased latency in decision-making.

Moreover, a custom data fabric can limit adaptability. As cybersecurity needs evolve and new threats emerge, data models and processing requirements must also change. An in-house team may find it difficult to quickly integrate new data sources, build real-time processing capabilities, or support new AI/ML models. A commercial data fabric, on the other hand, is continuously updated to accommodate new technologies, regulations, and security practices.

How will you handle integration with your current security stack and any future tools?
Effective security data fabrics (like Zscaler Data Fabric for Security) are designed to connect seamlessly with other security tools—like endpoint detection, vulnerability scanners, CMDBs, SIEMs, and compliance tools—enabling unified views of security data across platforms. A custom-built fabric often struggles to achieve such interoperability, given the proprietary formats and APIs of many security tools. Off-the-shelf solutions come with pre-built connectors and compatibility with a wide range of security systems, ensuring smoother integrations and faster deployments.

Custom solutions require ongoing development to maintain compatibility as third-party tools update their APIs, introduce new features, or adopt new data formats. The manual effort required to manage these integrations can slow down operations and increase the risk of security gaps.

Are you prepared for ongoing security and compliance risks?
Building a security data fabric introduces another layer of risk: securing the fabric itself. Data fabrics handle sensitive, business-critical information, making them an attractive target for cyberattacks. Without the robust, battle-tested security measures found in commercial solutions, a custom-built fabric may expose organizations to breaches or compliance violations. Implementing and maintaining security features—such as access controls, encryption, and compliance tracking—adds to the complexity and cost of building a custom fabric.

PART 2

The Zscaler Data Fabric for Security

How did the founders of Avalor—now part of Zscaler—approach building a data fabric differently?

The founding of Avalor, now part of Zscaler, was driven by the realization that security is fundamentally a data problem. Avalor's founders were data experts rather than cybersecurity specialists, which set them apart from the beginning. Their backgrounds in data integration and analytics, honed through previous work at Datorama (later acquired by Salesforce), uniquely positioned them to tackle one of cybersecurity's most persistent challenges: disconnected and siloed data. Drawing from their expertise in marketing data integration, the founders recognized a similar issue within security—massive amounts of disjointed data being generated without a cohesive way to integrate, analyze, and act on it.

The inspiration for Avalor stemmed from conversations with security leaders and practitioners who were struggling to handle overwhelming volumes of data from disparate tools. Organizations were using dozens of security solutions, from vulnerability scanners to identity management systems, all generating unique data sets that did not integrate easily. Security teams often lacked the time and expertise to manually stitch together insights across these tools, which created delays and inconsistencies in understanding risks, threats, and overall security posture. Avalor's founders saw this fragmented landscape as a perfect use case for a data fabric, which could serve as a unified platform to aggregate, normalize, and enrich security data from multiple sources.

Building on the concept of data fabrics that the founders had previously applied to marketing data, they aimed to create a solution tailored specifically for security teams. They focused on building a "data-first" platform—one that aggregated, de-duplicated, and normalized data without locking into a specific use case. This approach ensured that security data could be collected, analyzed, and acted upon with speed and accuracy. It also made the platform highly flexible, allowing organizations to adapt to evolving security needs, incorporate new data sources, and develop new applications on top of the existing data infrastructure.

A key innovation in Avalor's approach was the use of an entity-based model instead of traditional event logs. This shift enabled the fabric to present a clearer picture of security data, focusing on entities like users, devices, or IP addresses, and how they relate to each other. By leveraging this entity-driven structure, Avalor's data fabric could provide more meaningful insights, better context, and more accurate risk assessments, distinguishing it from traditional security tools like SIEMs.

Another critical element of Avalor's vision was the democratization of data. They believed that data should not be restricted to data scientists or security analysts alone but should be accessible and usable across the entire security organization.

To achieve this, Avalor's data fabric was designed to offer easy integration, advanced enrichment, and rapid visualization, making it a comprehensive platform for various security operations, including vulnerability management, threat detection, and compliance reporting.

The founders' commitment to flexibility, scalability, and adaptability in a data fabric framework attracted the attention of Zscaler, a leading cloud security provider. Recognizing the potential of Avalor's approach, Zscaler acquired the company to enhance its security capabilities with a robust data integration layer. This acquisition marked Zscaler's largest to date and underscored the growing importance of data fabrics in the cybersecurity landscape. Today, Avalor's original vision continues to drive Zscaler's security offerings, serving as a foundational component of its data architecture and powering multiple applications across the organization.

How does the Zscaler Data Fabric for Security work?

The Zscaler Data Fabric is built around five core components, each playing a critical role in data integration, transformation, and security management:

1. **Ingesting:** The first step involves ingesting data from various sources, such as cloud services, on-premise applications, and third-party security tools. The data fabric establishes connections with multiple data streams using APIs, batch uploads, or real-time streaming. This broad connectivity ensures that the data fabric can capture diverse data, ranging from user activity logs to vulnerability scans, providing a comprehensive view of security across an organization.
2. **Formatting:** Once ingested, data undergoes formatting to ensure consistency. This process standardizes different data formats (structured, semi-structured, or unstructured) into a common schema. For example, IP addresses, device identifiers, and log timestamps may come in varied formats depending on the source system. By converting these into a unified format, the Zscaler Data Fabric ensures that subsequent steps in the pipeline can operate smoothly, minimizing errors and discrepancies.
3. **Enriching:** After formatting, the data is enriched with additional context. This involves integrating metadata, correlating data points, and adding insights from other sources. For instance, a vulnerability detected in an endpoint may be enriched with information about the asset's criticality, ownership, or historical risk levels. This step enhances the quality of data and provides a more meaningful view, facilitating better decision-making.
4. **Resolving and deduplicating:** One of the major strengths of the Zscaler Data Fabric is its ability to resolve and deduplicate data across sources. This component aims to identify and merge duplicate records, such as the same IP address appearing in different logs, by using entity resolution algorithms. By ensuring that each entity is uniquely represented, this step helps reduce noise and improve data accuracy, enabling a clearer understanding of security events and asset status.

5. **Grouping:** The final step involves grouping related data points to form coherent entities or clusters. For example, multiple vulnerability alerts can be grouped into a single asset–based risk profile. This approach allows security teams to gain a more holistic view of related events, entities, or threats, facilitating more efficient analysis and faster incident response.

The Zscaler data fabric architecture offers a data–first approach, aggregating, de–duplicating, and normalizing data without focusing on a specific use–case initially. Once organized and prioritized, the data becomes actionable to allow customers to:

- **Easily include business logic:** After data processing, organizations can incorporate their specific business rules into the data fabric, ensuring that the platform aligns with unique operational needs.
- **Quickly build dashboards and reporting:** The data fabric supports the rapid creation of dashboards and reports, providing real–time visibility into security metrics and trends.
- **Seamless integrate workflow automation:** It integrates into existing workflows, ensuring that data insights can be acted upon without disrupting established processes.
- **Maintain lineage:** The data fabric maintains a complete history of data transformations, ensuring traceability and compliance.
- **Control who can see and change data:** With robust role–based access controls, the data fabric ensures that sensitive data is only accessible or modifiable by authorized users.
- **Capture changes over time:** The Zscaler Data Fabric for Security tracks changes in data, allowing security teams to analyze trends, spot anomalies, and ensure data integrity over time.

What makes the data model at the heart of the Zscaler Data Fabric for Security unique?

The Zscaler Data Fabric for Security employs a distinct data model designed to optimize flexibility, adaptability, and accuracy. Here's what sets it apart:

Entity–based architecture

Unlike traditional security data solutions that focus primarily on event logs, Zscaler's data model is built around entities. Entities represent assets, users, IP addresses, or other discrete elements within an organization's security landscape. This approach offers a more meaningful, relationship–based view of security data, enabling deeper insights and context–driven analysis. By focusing on entities rather than just raw events, the Zscaler Data Fabric enables better correlation, visualization, and risk assessment.

Opinionated but flexible

One of the key principles of the Zscaler data fabric model is that it is opinionated, meaning that it is built with specific principles and strong assumptions about how data should be structured, managed, and accessed. It emphasizes prescriptive structures, promoting a "convention over configuration" approach that prioritizes specific use cases.

Customizable and extensible schema

The Zscaler data fabric model is designed to be dynamic, with customizable schemas that can adapt to changing security needs. Organizations can define their own entities, attributes, and relationships, allowing the fabric to adjust as new data sources or security tools are integrated. This extensibility is critical in keeping the data model relevant and aligned with evolving security requirements.

Evolves over time

The semantic data layer in the Zscaler data fabric for security serves as a critical component in structuring and interpreting the vast amounts of security data it aggregates. The semantic data layer enables the separation between how data is stored and how it's accessed, allowing for a flexible model that can evolve with changing data structures and requirements. Within this layer, relationships between entities—such as assets, tickets, findings, and owners—are mapped, enabling users to query data meaningfully without needing to understand the backend storage. This abstraction is crucial for integrating diverse data sources and generating actionable insights across various security applications, like vulnerability management or compliance. By organizing data in a semantic layer, the Zscaler Data Fabric for Security creates a unified view of security information, enabling rapid adaptation to new data sources and evolving security requirements without extensive reconfigurations.

Seamless integration with other tools

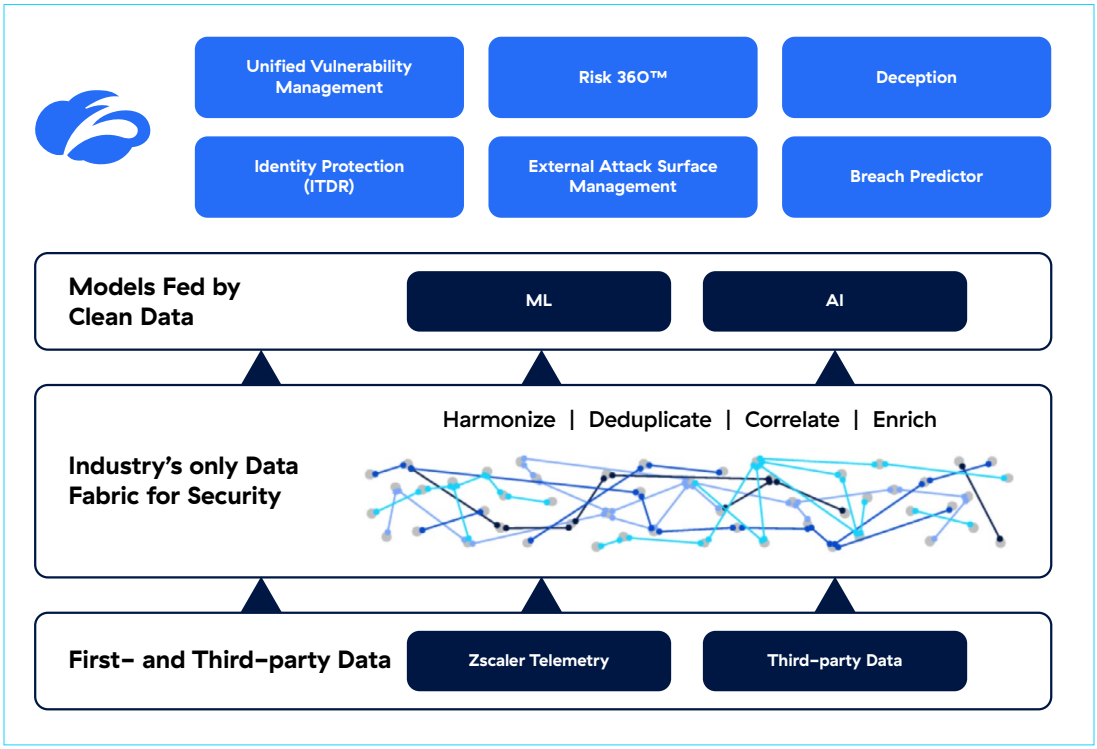
The Zscaler Data Fabric's model supports seamless integration with a variety of security tools, such as SIEMs, vulnerability management systems, and compliance tools. This is achieved through APIs, connectors, and pre-defined mapping for common data formats. The model also facilitates the exchange of entity-based insights across platforms, breaking down silos and enabling a unified security view.

Supports rapid application development

The flexible data model makes it easy to build new applications on top of the data fabric, leveraging the same foundational data architecture. For example, applications for unified vulnerability management, cyber risk quantification, or identity management quickly plug into the existing data structure. This speeds up development time and enables quicker deployment of security solutions that are both tailored and scalable.

What solutions does the Zscaler Data Fabric for Security power?

Today, the Zscaler Data Fabric for Security serves as the foundational architecture for Unified Vulnerability Management (UVM) and Risk360. In the near future, it will power additional solutions across the Zscaler Risk Management portfolio.



The rich integrations available within the fabric additional Zscaler solutions to connect back into the fabric and provide rich feedback loops between products. For example, Zscaler EASM feeds external attack surface findings and asset information to UVM for risk prioritization and to Risk360 for cyber risk quantification. Zscaler DSPM offers data findings to UVM to use as a risk factor, and UVM sends back remediation assignments for elements that need to be corrected. And, when UVM or Risk360 uncovers a risk insight that necessitates an access control change, they leverage the connectors within the data fabric to send policy adjustment details to Zscaler ZIA and ZPA.

Conclusion

The concept of a data fabric has become increasingly vital in today's complex data environments, offering organizations a way to unify, manage, and leverage data across diverse sources and platforms. With the exponential growth of data generated by businesses and the rising need for real-time decision-making, traditional data solutions—like data warehouses and data lakes—fall short in achieving seamless integration, governance, and accessibility. A data fabric's metadata-driven, real-time integration addresses these gaps by providing a single, consistent layer for data access, governance, and analytics, making it indispensable for strategic decision-making.

In the cybersecurity domain, the fragmented nature of data remains one of the most significant challenges. Security teams often struggle with disparate data streams from a multitude of tools, leading to a lack of visibility and slow response times. The Zscaler Data Fabric aims to resolve this by serving as a unifying layer that not only aggregates but also enriches and correlates data across the security ecosystem. By doing so, it creates a "single source of truth," enabling faster threat detection, better risk prioritization, and real-time response to emerging threats.

Zscaler's approach to building a security-focused data fabric, rooted in an entity-driven architecture, sets it apart from traditional SIEMs and other security tools. It prioritizes data quality, context, and interoperability, enabling deeper insights and more accurate risk assessments. This makes it a foundational component of a more adaptive, efficient, and scalable security strategy.

Ultimately, as organizations look to optimize their data management and security operations, the adoption of data fabrics—like Zscaler's—becomes not just beneficial, but necessary. By breaking down data silos, enhancing analytics capabilities, and supporting real-time decision-making, a data fabric empowers security teams to protect their organizations more effectively in an increasingly data-centric world.

Glossary of Terms

- **Data Fabric:** An architecture that enables seamless data integration, management, and access across diverse systems and environments, providing a unified view of data for better analytics, governance, and decision-making.
- **Data Fabric for Security:** A Data Fabric designed specifically to support entities relevant in addressing security use cases.
- **Entity:** An entity refers to any distinct unit that possesses an identifiable existence within a system, process, or database. Entities are fundamental components in system design, databases, and security frameworks, as they define the various elements that need to be managed, monitored, or protected. Entities can represent various components, such as servers, routers, programs, platforms, users, owners, etc.
- **Entity resolution:** The process of identifying and merging records across databases that represent the same entity, despite variations in attributes.
- **Outegration:** A Zscaler term for outbound connectors/integrations to other systems, typically referring to ticketing and project management systems but can also be data lakes, SIEMs, or anything an organization needs to send operational data out to.
- **Semantic Data Layer:** An abstraction layer that provides a unified, business-oriented view of data by defining common terminologies and relationships, facilitating better understanding and usage of data across applications.
- **ETL (Extract, Transform, Load):** A data integration process that extracts data from sources, transforms it into a desired format, and loads it into a target system or database for analysis.
- **Open Security Graph™:** The representation of the relationships between different entities, made possible by the Zscaler Data Fabric for Security and its extendable, customizable data model.

+1 408.533.0288 Zscaler, Inc. (HQ) • 120 Holger Way • San Jose, CA 95134 zscaler.com