ホワイト ペーパー

Data Fabric For Security — Zscaler のソリューションで セキュリティ データの課題を解消

目次

はじめに

デジタル トランスフォーメーションがビジネスを大きく変化させている現代において、データは資産であると同時に課題でもあります。クラウド コンピューティングや IoT、AI といった技術が急速に採用されたことで、データの量、種類、速度は爆発的に増加しています。この変化に伴って複雑なマルチクラウド環境やハイブリッド環境が生み出され、データを効果的に統合、管理、活用することがますます難しくなってきています。データを保存して分析するように設計されたデータ ウェアハウスやデータ レイクなどの従来のデータ管理ソリューションでは、シームレスな統合、リアルタイムの分析、包括的なガバナンス、インテリジェントを十分に提供できません。

このようなより高度な要件に対応するには、データ ファブリックが必要です。データ ファブリックは、異なるデータ ソースを統合するアーキテクチャー フレームワークであり、さまざまな環境にわたって一貫したリアルタイムのデータ管理レイヤーを提供します。アクセスの簡素化、ガバナンスの強化、スピーディーな分析のサポートで戦略的優位性を確立し、より迅速で情報に基づいた意思決定を可能にします。実際、Gartner は Top Strategic Technology Trends for 2022 のレポートで、データ ファブリックの導入によりデータ管理の労力が最大 70% 削減され、価値実現までの時間が短縮されると推定しています。

サイバーセキュリティの領域では、リアルタイムのインテリジェンスと迅速な脅威検出が非常に重要であるため、データ統合の課題がさらに顕著になっています。さまざまなツールが生成する大量のデータは統合されていないため、可視性が断片的になり、対応に遅れが生じます。これは多くのセキュリティ部門が直面している問題です。Zscaler は、革新的なアプローチでデータ ファブリック アーキテクチャーに取り組むことでこの問題を解消します。データを正規化して強化し、関連付けるセキュリティ重視のソリューションにより、優れたインサイトの確保と、脅威に対するより迅速かつ効果的な対応が可能になります。

このホワイト ペーパーでは、データ ファブリックの基礎、セキュリティにおけるデータ ファブリックの役割、セキュリティ オペレーションを再定義する Zscaler Data Fabric for Security の仕組みについて説明します。Zscaler Data Fabric for Security は、リアルタイムの意思決定の推進からリスクの優先順位付けの改善までを可能にすることで、現代のサイバーセキュリティ戦略におけるデータ ファブリックの可能性を実証します。

パート 1

データ ファブリックとは

データ ファブリックの概要

データ ファブリックはマルチクラウドやハイブリッド、複雑なデータ エコシステムでデータを管理するための包括的なソリューションです。データ アクセスを簡素化すると同時に、ガバナンスを強化し、リアルタイムの分析をサポートするため、データを戦略的資産として活用したい組織にとっては不可欠なフレームワークとなっています。データ ファブリックにより、運用の合理化と意思決定の改善が可能になり、最終的により優れた成果を得られるようになります。

しかし、データ ファブリックとは正確には何なのでしょうか？簡潔にまとめると、データ ファブリックは、多様で複雑な環境間でのデータの統合、管理、配信を可能にするアーキテクチャー フレームワークです。オンプレミス、クラウド、ハイブリッド環境などのさまざまなデータ ソースをつなぎ合わせ、データ アクセスとガバナンスが統一された一貫したレイヤーを提供します。このアプローチはさまざまなソースからの膨大な量の構造化データ、半構造化データ、非構造化データを処理するためにデータ管理が複雑になる問題に対処します。

データ ファブリックは、情報技術がもたらす多くの成果を推進するために利用されていますが、次のような基盤となる共通の機能を備えています。

データ アクセスの統合：データ ファブリックは、異なるプラットフォームと形式のデータを統合し、データの物理的な保存場所にかかわらずアクセスできる単一の環境を構築します。基盤となるインフラは抽象化されるため、ユーザーやアプリケーションはデータの実際の保存場所を知らなくても必要なデータにアクセスして操作できます。

メタデータを活用したアプローチ：データ ファブリックは、データ管理を自動化するためにメタデータ (データに関するデータ) を活用します。このアプローチにより、データ検出、ガバナンス、アクセスが効率的かつ容易になります。データ ファブリックはデータ資産間の関係を理解することで、動的なデータ オーケストレーションを可能にし、適切なデータを適切なタイミングで利用できるようにします。

データ ガバナンスとセキュリティ：データ ファブリックは、データを取り巻くあらゆる環境で一貫したガバナンスとセキュリティ ポリシーを施行するため、データのソースや場所に関係なく、データのプライバシーと規制の要件が確実に満たされます。このフレームワークでは、ポリシーの一元管理が可能ですが、必要に応じてローカライズされた制御も適用できます。

リアルタイムのデータ アクセスと分析：データ ファブリックは、リアルタイムまたはほぼリアルタイムのデータ処理と分析をサポートするように設計されています。最新の情報を利用して意思決定を行う現代の組織にとっては必須といえる機能です。データにリアルタイムでアクセスして分析する機能により、変化する市場状況や顧客の行動、運用上の課題に迅速に対応できます。

セキュリティとリスクを適切に管理するためにデータ ファブリックが必要な理由

現代において、サイバーセキュリティはデータと同義です。セキュリティ ソリューションは、個人と組織の両方のデータを保護することを目的としていますが、その過程で膨大な量のデータが生成されます。しかし、こうしたデータは関連付けられていないばかりか、効果的にデータを統合して解釈するためのリソースや専門知識も不足しているため、セキュリティ リーダーの大きな課題となっています。

セキュリティ部門がより多くのデータを求めていたのは、それほど昔のことではありません。初期のツールは、基本的なセキュリティ指標と組織のセキュリティ態勢に関するレポートを提供していましたが、リスク インテリジェンスとリアルタイムのレポートに対する需要が高まるにつれ、CISO はよりデータに基づいたインサイトを提供するソリューションを導入して、効果的なセキュリティ対策を確保したいと考えるようになりました。

市場は熱狂的にこれに反応したため、MDR から脆弱性スキャナー、脅威インテリジェンス プラットフォームまで、さまざまなセキュリティ ソリューションが過去 20 年の間に登場しました。現在、多数の大企業がセキュリティ スタックに 30 種類以上のツールを導入しており、各ツールから独自のデータが生成されています。しかし、こうしたデータの目的はツールごとに異なるため、他のソリューションからのデータと統合されたり、重複排除が行われたりすることはほとんどなく、結果的に分断された情報が膨大になります。通常、こうしたデータを統合し、セキュリティの性能を測定するうえで苦戦するのは、セキュリティの専門家ではなく、データの専門家なのです。

残念ながら、セキュリティ部門はデータから価値とコンテキストを抽出するのに苦労しており、これらのツールやプログラムが期待どおりに機能していないことは明らかです。さらに、セキュリティ部門が保有するデータはサイロ化しており、セキュリティの脅威と脆弱性を統一して包括的に把握することができません。

データ ファブリック アーキテクチャーを実装することであらゆるデータ ソースを統合し、多数のユース ケースに対応するオープンでスケーラブルな基盤を構築できます。この柔軟なシステムは、組織全体から生成される複数のデータ ポイントに対応し、信頼性と実用性に優れたインサイトを提供します。

信頼できるソース：サイバーセキュリティ部門がデータ ファブリックを必要とする主な理由の一つは、さまざまなセキュリティ ツールやベンダーによってデータが断片的になる問題を解消するためです。単一のシステム (営業の Salesforce など) で全体像を把握できる他の分野とは異なり、サイバーセキュリティは多数の専用ツールで運用され、それぞれが脆弱性管理や資産管理、セキュリティ情報とイベント管理 (SIEM) などセキュリティのさまざまな側面に焦点を当てています。これらのツールはデータを別々の環境に保存している場合が多いため、セキュリティ環境を完全かつ正確に把握することは難しくなっています。

データ ファブリックは、こうしたすべてのツールから生成されたデータを統合する統合レイヤーとして機能し、「信頼できる唯一の情報源 (Single Source of Truth)」を構築します。さまざまなシステムからのデータを変換して関連付けるため、セキュリティ環境全体を検索して分析し、インサイトを導き出すというプロセスが簡素化されます。データ ファブリックが提供する統合されたビューにより、サイバーセキュリティの専門家はユーザー、資産、脆弱性など、さまざまなデータ ソースとエンティティー間の関係を理解できるようになり、さらに効果的な脅威の検知と対応を確保できます。

リアルタイムの意思決定の強化：日々進化するサイバー脅威のスピードに追いつくには、リアルタイムの意思決定が必要です。しかし、統一されたデータ アーキテクチャーを実装していない場合、セキュリティ部門は膨大な時間を費やしてさまざまなソースからのデータを関連付けなければならず、結果として対応までに時間がかかったり、脅威が急増したりする可能性があります。データ ファブリックは、データをリアルタイムで正規化および重複排除することで、意思決定を加速させます。セキュリティ データが包括的に統合、強化された単一のビューにより、複数のシステムからの情報を手動で関連付けることなく、ネットワーク全体の状況をすばやく理解できます。

さらに、データ ファブリックは、強化とエンティティー解決の自動化もサポートします。たとえば、さまざまなソースからのデータ ポイントを自動的に解決することで、特定の資産や脆弱性に関する情報が正しく集約され、相互参照されるようにすることができます。これにより、データが不完全なために重大な脆弱性を見逃すなどといった問題が発生するリスクが低くなります。データ ファブリックでさまざまなシステムからのデータを解決し、関連付ければ、セキュリティ態勢をより明確かつ正確に理解できるようになり、情報に基づいた意思決定をさらに迅速に行えるようになります。

リスクの優先順位付けの精度向上：サイバーセキュリティでは、脅威を適切なタイミングで正確に検出することが重要です。しかし、従来の SIEM システムでは主に脅威検出のために生データを集約するため、包括的なセキュリティ分析に必要なコンテキスト化されたインサイトは提供されません。データ ファブリックは、生データを実用的なインテリジェンスに変換することで、脅威検出を強化します。脅威インテリジェンス フィードなどの外部ソース、ユーザーや資産のデータなどの内部ソースと関連付けることでデータを強化し、潜在的な脅威をより包括的に把握できるようにします。

また、データ ファブリックでは、より効果的なデータの相互相関が可能になるため、他の方法では見落とされやすいパターンや関係性を特定できます。

たとえば、ユーザーのアクティビティー データを既知の脆弱性や外部の脅威インテリジェンスと関連付けて、潜在的なリスクをより詳細に理解することができます。個々のデータ ポイントを単独で分析しても検出できない可能性のある、複雑なマルチベクトル攻撃を特定するためには、このレベルの関連付けが不可欠です。

データ ファブリックとデータ ウェアハウスとデータ レイクの違い

データ ファブリック、データ レイク、データ ウェアハウスは、それぞれ組織内のデータ管理において独自の役割を果たします。いずれも広範なデータ アーキテクチャーの一部ですが、その設計、目的、扱うデータの種類は大きく異なります。ここでは、これら 3 つの違いを詳しく見ていきます。

データ ファブリック：あらゆるソースでの包括的かつリアルタイムの統合

データ ファブリックは、オンプレミス、クラウド、エッジのいずれであっても、複数のソースから生成されたデータを統一し、統合するように設計された包括的なデータ アーキテクチャーです。統一されたアクセス レイヤーを提供することで、組織全体のデータの一貫したリアルタイム ビューが可能になります。データ ファブリックは、データ レイクやデータ ウェアハウスのようにデータ ストレージだけに焦点を当てるわけではなく、シームレスなデータ統合、ガバナンス、検出、オーケストレーションもサポートしています。このアプローチにより、データのサイロ化が減り、部門間のアクセシビリティーと意思決定の質が向上します。

データ ファブリックは構造化データ、半構造化データ、非構造化データを統合し、データの専門家だけでなく、幅広いユーザーがリアルタイムでアクセスできるようにします。

こうしたデータの多くはデータ ウェアハウスやデータ レイクに保存されます。一方、データ ファブリックはデータを取得すると、高度な分析、機械学習、自動化機能を駆使して、さまざまなソースから生成されたデータを迅速に処理し、コンテキスト化します。このアーキテクチャーは、アクセス制御ポリシーとガバナンス ポリシーを施行することで、複雑なプロセスのサポート、データ セキュリティの強化、コンプライアンスの確保を可能にします。ただし、データ ファブリックを単独で実装するのは複雑で、専門的なスキル、綿密に計画された統合戦略、多額の投資が必要になります。ポイント：必要な作業はすべて Zscaler で行われるため、お客様側での作業は一切発生しません。

データ ウェアハウス：事前定義された分析用の構造化データ
データ ウェアハウスは、クレンジング、変換、および事前定義済みスキーマに編成された構造化データと精製されたデータを保存するように設計されています。高速の分析処理に対応し、ビジネス インテリジェンス、レポート作成、トレンド分析に使用されます。抽出、変換、ロード (ETL) プロセスによってさまざまなソースから生成されたデータを統合し、一貫性と精度を確保します。

データ ウェアハウスは、処理されたデータから信頼性の高いインサイトを迅速に提供するため、運用上の意思決定に欠かせません。複雑な SQL クエリーの実行、管理部門や事業部門向けのレポート作成に適しています。ただし、非構造化データを保存したり、生データを使った実験を処理したりする柔軟性には欠けており、新たな分析や探索的プロセスでの有用性は限られます。

データ レイク：生データ ストレージの一元化
データ レイクは、構造化データ、半構造化データ、非構造化データなど大量の生データを元の形式で保存する一元化されたリポジトリーです。ログ ファイルや IoT センサー データから、音声、動画、ソーシャル メディア フィードまで、膨大な量のデータを処理するように設計されています。データの処理と分析の準備が整うまで、柔軟で費用対効果の高いストレージを提供することが主な機能です。

データ レイクは、データ ウェアハウスのように特定のスキーマで事前設定されていないため、実験や機械学習、AI などの高度な分析に適しています。多くの場合、データ レイクに含まれる多様なデータ タイプを管理および処理するために、Hadoop や Apache Spark、NoSQL データベースなどの技術が必要です。しかし、データ レイクはその生データという性質上、メタデータ、品質管理、ガバナンスの対策で適切に管理されていないと、すぐに「データの沼」になってしまいます。また、生データのクエリーは複雑で時間がかかる可能性があるため、データ レイクから価値を抽出するための専門的なスキルも必要になります。

データ品質階層

さまざまなデータ アーキテクチャーを検討する際には、その中のデータの品質と用途を考慮することが重要です。これを検討する方法の一つとして挙げられるのが、データ品質階層です。「ゴールド」、「シルバー」、「ブロンズ」のデータ レイヤーはデータ品質、粒度、データ アーキテクチャー内で使用するための準備状況の階層を表します。これらのレイヤーは、データ ワークフローの管理と最適化に不可欠な要素であり、データから効率的に価値を引き出せるようにします。

ゴールド レイヤーは、最も粒度が高く高品質で、分析に対応したデータで構成されています。このレイヤーのデータは標準化された形式とメタデータによって、クレンジング、定義、調整などがすべて行われます。通常はシルバー レイヤーから派生し、重要な意思決定、高度な分析、ビジネス インテリジェンスに使用されます。ここでのデータは多くの場合、集計または要約され、精度、一貫性、使いやすさの最高基準を満たします。ダッシュボード、レポート、機械学習モデルなど、精度が重要なケースに適しています。

シルバー レイヤーは、データ処理の中間段階を表します。このレイヤーには、初期のクレンジング、変換、検証を経て部分的に調整されたデータが含まれています。シルバー データは完全に洗練されているわけではありませんが、生データよりも使いやすく、探索的分析や予備的なレポート作成に最適です。通常、ある程度の粒度を保持しているデータが含まれているため、ゴールド レイヤーに昇格する前に、より詳細な調査とトラブルシューティングが可能になります。シルバー データは橋渡しの役割を担い、生データを完全な集約なしでより使いやすい形式に変換します。

ブロンズ レイヤーは、さまざまなソースから取り込まれた未処理の生データで構成されています。多くの場合、リアルタイムまたはバッチ ロードで取り込まれます。このデータは元の形式で保存されるため、最も粒度が低い状態です。他のすべての処理の基盤レイヤーとして機能し、収集されたデータの完全な記録を提供します。ブロンズ データはすぐに分析に使用できるわけではありませんが、履歴分析、異常検出、デバッグには不可欠です。

以上の階層化された構造により、データ管理の柔軟性、スケーラビリティー、ガバナンスが向上します。

	データ ファブリック	データ ウェアハウス	データ レイク
目的	データの統合、ガバナンス、アクセシビリティーをすべてのユーザーに提供し、信頼できる唯一の情報源を作成し、複数のユース ケースに対するデータのサイロ化を緩和します。	レポート作成やビジネス インテリジェンスのための履歴データの高速分析をサポートします。	生データ用の大規模で費用対効果の高いリポジトリーを提供し、探索や高度な分析に最適化します。
サポートされているデータの種類	すべてのソースの構造化データ、半構造化データ、非構造化データ。	事前定義されたスキーマのある、構造化と処理が行われたデータのみ。	事前定義されたスキーマのない、すべての形式の生データ。
データ品質階層	ゴールド	シルバー	ブロンズ

データ ファブリックと他のセキュリティ アーキテクチャー (SIEM など) の違い

セキュリティ データ ファブリックおよびセキュリティ情報とイベント管理 (SIEM) は、サイバーセキュリティ エコシステム内で異なる役割を果たし、データの処理に対するアプローチも異なります。主な違いは、その目的とアーキテクチャーにあります。

データ ストレージ vs. データ最適化
SIEM とセキュリティ データ ファブリックの最も重要な違いの一つは、データ ストレージへのアプローチです。従来の SIEM は、大量のログ データを集約して保存するという考え方に基づいて構築されています。この大規模なデータ ストレージは、処理および保存されたデータの量に基づいて課金するため、SIEM ソリューションにかかるコストが増加する可能性があります。収集されるデータが多いほど、消費コストとライセンス コストが高くなるこのモデルにより、SIEM ベンダーはできるだけ多くのデータを保存しようとします。

対照的に、データ ファブリックは、データを効率的に最適化および管理するように設計されています。情報をさまざまな宛先にルーティングする前に、正規化、強化、重複排除してデータ量を削減します。データ ファブリックにより、組織は低コストのストレージにデータを送信できるため、ストレージの費用をさらに削減できます。この最適化により、データ品質が向上するだけでなく、組織全体でさまざまな種類のセキュリティ データを処理する際の柔軟性も向上します。ここでの主な課題は、データ量の増加が SIEM ベンダーの収益を牽引するため、こうしたベンダーにはデータ量の削減を優先する理由がほとんどないということです。

イベント vs. エンティティー
SIEM は基本的にイベント ログを活用します。ログ データをほぼリアルタイムで収集、集約、分析し、異常やセキュリティ脅威を特定します。各イベント ログは、ログイン試行、ファイル アクセス、システム エラーなどの個別のアクションやインシデントを表します。SIEM の主な目的は、これらのログのパターンを調べることで潜在的な脅威を検出することです。そのため、SIEM は、迅速なログ分析とイベントの関連付けが必要な場合に適しています。また、個々のイベントを把握し、それらを組み立ててアラートを生成することに常に焦点を当てているため、リアルタイムの監視、インシデント対応、フォレンジック調査に特に効果的です。

一方、データ ファブリックはエンティティーを活用します。データを単発のイベントとして扱うのではなく、関連する情報を結び付けて、デバイス、ユーザー、アプリケーション、資産などのエンティティーを包括的に理解できるようにします。このアプローチにより、データを収集するだけでなく、主なエンティティーを中心に強化、正規化、関連付けを行う、より包括的なビューが可能になります。たとえば、データ ファブリックは、複数のソースからのデータをマージして、資産やユーザー、脆弱性に関する統一された単一のビューを提供します。エンティティーとその関係性に焦点を当てることで、セキュリティ リスク、資産管理、コンプライアンスに関するより深いインサイトを得ます。

SIEM のイベント ログ中心のアプローチは、特定のインシデントの検出と対応に理想的ですが、データ ファブリックのエンティティーに基づく視点は、より広範なコンテキストを提供するため、長期的なセキュリティと運用のニーズに対応できます。こうした差別化により、データ ファブリックはより適応性が高く、包括的なセキュリティ戦略の重要な要素となっています。

独自のデータ形式 vs. 相互運用性

もう一つの大きな違いは、データ形式と統合です。SIEM は通常、独自のデータ形式に基づいて構築されているため、相互運用性に対処できない場合があります。その結果、SIEM では他のツールとのスムーズな統合や、競合する SIEM 製品のデータ形式の管理で問題が発生する可能性があります。データ形式を標準化する取り組みは行われているものの (Open Cybersecurity Schema Framework など)、完全な互換性の達成にはいまだ至っていません。

一方、セキュリティ データ ファブリックは、さまざまなデータ形式や宛先と連携して、相互運用できるように明示的に設計されています。この機能は、複数のデータ形式に対する柔軟性とサポートが不可欠である、マルチ SIEM やセキュリティ データ レイク戦略を実装する組織にとって非常に重要です。このファブリックでは異なるデータ ソースを統合できるため、複雑なセキュリティ上の課題への対応に適しており、全体像を把握できます。これは、SIEM では実現できません。

ベンダーへの依存の増大 vs. すべてに接続可能な機能

セキュリティ データ ファブリックを使用する主な目的は、SIEM ベンダーへの依存を回避することです。SIEM はその複雑な統合により、組織のセキュリティ運用に深く組み込まれることが多く、ベンダーの切り替えにはコストも時間もかかります。ベンダーの市場支配が維持されるように、SIEM 製品は競合他社への移行や他のシステムとの統合が簡単にはできないように設計されている場合が少なくありません。

これに対し、データ ファブリックはセキュリティ データに依存しないように構築されており、さまざまなツールやソリューション間でのシームレスなデータ移動を可能にします。この柔軟性により、組織はソリューションをより簡単に切り替え、特定の SIEM のエコシステムに限定されることなく、利用可能な最高の技術を活用できます。データの収集と取り込みを脅威分析、検出、フォレンジックという主要機能から切り離すことで、柔軟性を維持しながら、全体的なセキュリティ アーキテクチャーを強化します。

SIEM とデータ ファブリックは重複する機能を持っているように見えるかもしれませんが、競合する技術ではなく、それぞれが補完的な役割を果たすことを理解する必要があります。SIEM は、リアルタイムの監視、アラート生成、フォレンジック分析など、実行に最適化された主要機能を備えています。データ ファブリックは、データの準備、統合、最適化に優れており、すべてのセキュリティ データの「信頼できる唯一の情報源」として機能します。

	データ ファブリック	SIEM
目的	すべてのユーザーがデータの統合、ガバナンス、アクセシビリティーを利用できるようにすることでデータのサイロ化を減らし、リアルタイムのインサイトを提供します。	個々のセキュリティ イベントをキャプチャーし、それらを組み立ててアラートを生成します。
サポートされているデータの種類	すべてのソースの構造化データ、半構造化データ、非構造化データ。	リアルタイムの監視、アラート、クエリー分析のためのイベント ログ。
データ品質階層	ゴールド	シルバー

セキュリティ データ ファブリックを組織で構築する場合に考慮すべきポイント

データ インフラを完全に制御したい組織にとって、独自に構築したセキュリティ データ ファブリックは魅力的に思えるかもしれません。しかし、そのような取り組みは複雑で、コストがかかるばかりか、継続的な要求も発生するため、リスクが高く、欠陥のある戦略になる可能性があります。ここでは、独自のセキュリティ データ ファブリックを構築しようとすると、大きな課題が生じかねない理由を説明します。

社内に必要な専門知識はあるか?

セキュリティ データ ファブリックの開発は、一般的なデータ パイプラインの作成やデータ ウェアハウスの実装とは異なります。データ ファブリックは、構造化、半構造化、非構造化など、複数のソースから生成された膨大な量のデータを統合、正規化、関連付けるように設計された多層アーキテクチャーです。このようなシステムを構築するには、データ統合、ガバナンス、セキュリティ、機械学習に関する広範な専門知識が必要です。また、さまざまなデータ形式、セキュリティ プロトコル、API を詳細に理解する必要もあります。

このレベルの統合と機能を実現することは簡単ではありません。これには、データの取り込みやエンティティーの解決から、強化や重複排除まで、あらゆるレベルでのカスタム エンジニアリングが求められます。高度なデータ サイエンス部門のある組織でさえ、さまざまなユース ケースや環境に効果的に拡張できるファブリックを構築するための専門知識を十分に備えていません。

構築と維持に多額の費用がかかることを理解しているか?

セキュリティ データ ファブリックをゼロから構築するには、技術と人員の両方に多額の先行投資が必要です。組織はインフラ、高度なソフトウェア、データ エンジニア、アーキテクト、サイバーセキュリティの専門家などの専門スタッフの採用 (またはトレーニング) に資金を割り当てる必要があります。

最初の構築後も、引き続きコストはかかります。データ ソースやデータ形式は頻繁に変更されるため、コネクター、パーサー、エンティティー モデルの定期的な更新、メンテナンス、最適化が必要です。新しいセキュリティ ツール、規制、データ標準にも継続的な調整が求められます。独自に構築したファブリックを適切なリソースがない状態で維持しようとすると、コストも時間もさらに必要になるため、技術的負債や敏捷性が低下する可能性があります。

スケーラビリティーと柔軟性をどのように管理するか?

データ ファブリック ソリューションでは、データの量と機能両方の観点から拡張できるように設計する必要があります。既製のデータ ファブリックは通常、複雑なデータ変換、大量の取り込み、ペタバイト規模のデータにわたる迅速な分析を処理するように構築されています。このレベルのスケーラビリティーを内部で確保しようとすると、パフォーマンスのボトルネックが発生し、処理が非効率的になる可能性があります。独自に構築したソリューションは、データ量が増加するとパフォーマンスを維持できなくなる場合があり、結果的にインサイトをタイムリーに入手できなかったり、意思決定に遅延が発生したりします。

さらに、独自のデータ ファブリックでは適応性が制限される可能性があります。サイバーセキュリティのニーズが進化し、新たな脅威が出現するにつれて、データ モデルと処理要件も変化する必要がありますが、社内のセキュリティ部門では、新しいデータ ソースの迅速な統合、リアルタイムの処理機能の構築、新しい AI/ML モデルのサポートなどには対処できない場合があります。一方、商用のデータ ファブリックは、新しい技術、規制、セキュリティ プラクティスに対応するために継続的に更新されます。

現在のセキュリティ スタックや将来のツールとの統合をどのように処理するか?

効果的なセキュリティ データ ファブリック (Zscaler Data Fabric for Security など) は、エンドポイント検出、脆弱性スキャナー、CMDB、SIEM、コンプライアンス ツールなどの他のセキュリティ ツールとシームレスに接続するように設計されており、プラットフォーム間でセキュリティ データを統合して表示できます。しかし、各セキュリティ ツールの形式と API を考えた場合、組織独自のファブリックではこのような相互運用性を確保できないケースがほとんどです。既製のソリューションには、事前に構築されたコネクターとさまざまなセキュリティ システムとの互換性が備わっているため、よりスムーズな統合と迅速な展開が可能になります。

組織独自のソリューションでは、サードパーティー ツールが API を更新したり、新機能や新しいデータ形式を導入するたびに、互換性を維持するために継続的な開発が必要になります。これらの統合の管理に必要な手作業は、運用を遅らせ、セキュリティ ギャップのリスクを高める可能性があります。

セキュリティとコンプライアンスの継続的なリスクに対する準備は万全か?

セキュリティ データ ファブリックを構築すると、ファブリック自体の保護という別の課題が発生します。データ ファブリックは、機密性の高いビジネスクリティカルな情報を処理するため、サイバー攻撃の格好の標的となっています。商用ソリューションのような堅牢で実績のあるセキュリティ対策がなければ、独自に構築したファブリックは組織を侵害やコンプライアンス違反にさらすおそれがあります。また、アクセス制御、暗号化、コンプライアンスの追跡などのセキュリティ機能を実装して維持しようとすると、複雑さとコストが増大します。

パート 2

Zscaler Data Fabric for Security

データ ファブリックの構築に対する Avalor (Zscaler 傘下) 独自のアプローチ

現在は Zscaler の傘下にある Avalor は、セキュリティは基本的にデータの問題であるという認識を基に設立されました。Avalor の創業者はサイバーセキュリティではなくデータの専門家であり、それが当初から他社とは一線を画していました。創業者は、Datorama (後に Salesforce が買収) での以前の仕事を通じて磨かれたデータ統合と分析のバックグラウンドを持ち、サイバーセキュリティの最も根強い課題の一つであるデータの分断とサイロ化に取り組むための独自の立場を確立しました。また、マーケティング データ統合の専門知識を活かして、セキュリティにおいても同様の問題があることを認識しました。つまり、統合、分析、対応するための一貫した方法がないまま、断片化された大量のデータが生成されていたのです。

Avalor のインスピレーションは、異なるツールから生成される圧倒的な量のデータの処理に苦労しているセキュリティ リーダーや実務担当者との会話から生まれました。多くの組織が脆弱性スキャナーやアイデンティティー管理システムなどのセキュリティ ソリューションを多数導入していましたが、いずれも簡単に統合できない独自のデータセットを生成していました。セキュリティ部門には、これらのツール間のインサイトを手動でまとめる時間も専門知識もなく、結果としてリスク、脅威、全体的なセキュリティ態勢の理解に後れや矛盾が生じていました。Avalor の創業者はこの断片化された状況こそが、複数のソースから生成されたセキュリティ データを集約、正規化、強化するための統合プラットフォームとして機能するデータ ファブリックの完璧なユース ケースだと考えました。

そして、以前マーケティング データに適用していたデータ ファブリックの概念に基づいて、セキュリティ部門に特化したソリューションを開発することを目指しました。同社は、特定のユース ケースに縛られることなく、データを集約し、重複を排除しながら、正規化する「データファースト」のプラットフォームの構築に注力しました。このアプローチにより、セキュリティ データを迅速かつ正確に収集、分析、処理できるようになったのです。また、プラットフォームの柔軟性も高まったことで、組織は進化するセキュリティ ニーズに適応しながら、新しいデータ ソースを組み込み、既存のデータ インフラ上に新しいアプリケーションを開発できるようになりました。

Avalor と他社との大きな違いは、従来のイベント ログではなく、エンティティーベースのモデルを採用したことです。この新しいモデルに移行したことで、ファブリックはユーザー、デバイス、IP アドレスなどのエンティティーと、それらが相互にどのように関連しているかに焦点を当て、セキュリティ データをより明確かつ包括的に表示できるようになりました。このエンティティーを活用した構造により、Avalor のデータ ファブリックは、SIEM のような従来のセキュリティ ツールとは一線を画し、より有意義なインサイト、より優れたコンテキスト、より正確なリスク評価を提供しています。

Avalor のもう一つの重要なビジョンとなっていたのが、データの民主化でした。同社は、データはデータ サイエンティストやセキュリティ アナリストだけに限定されるべきではなく、セキュリティ組織全体でアクセスと使用が可能であるべきだと考えていました。

これを実現するために、Avalor のデータ ファブリックは、簡単な統合、高度な強化、迅速な視覚化を提供するように設計されています。そして、脆弱性管理、脅威検出、コンプライアンス レポートなど、さまざまなセキュリティ運用のための包括的なプラットフォームとなっています。

データ ファブリック フレームワークの柔軟性、スケーラビリティー、適応性に対する Avalor の取り組みは、クラウド セキュリティを牽引する Zscaler の注目を集めました。Zscaler は Avalor のアプローチの可能性を認識し、堅牢なデータ統合レイヤーでセキュリティ機能を強化するために同社を買収しました。この買収は Zscaler にとってこれまでで最大のものであり、サイバーセキュリティ環境におけるデータ ファブリックの重要性が高まっていることを強調しています。Avalor の当初のビジョンは現在でも、Zscaler のセキュリティ製品を推進し続けています。データ アーキテクチャーの基盤となる要素として機能し、組織の複数のアプリケーションを強化しています。

Zscaler Data Fabric for Security の仕組み

Zscaler Data Fabric は、5 つの主要なコンポーネントを中心に構築されており、それぞれがデータ統合、トランスフォーメーション、セキュリティ管理において重要な役割を果たします。

1. **取り込み：**最初のステップでは、クラウド サービス、オンプレミス アプリケーション、サードパーティーのセキュリティ ツールなどのさまざまなソースからデータを取り込む必要があります。データ ファブリックは、API やバッチ アップロード、リアルタイムのストリーミングを使用して、複数のデータ ストリームとの接続を確立します。この広範な接続により、データ ファブリックはユーザー アクティビティー ログから脆弱性スキャンまで、さまざまなデータを取得し、組織全体のセキュリティを包括的に可視化します。

2. **形式化:**取り込まれたデータは、一貫性を確保するために形式化されます。このプロセスでは、さまざまなデータ形式 (構造化、半構造化、または非構造化) を共通のスキーマに標準化します。たとえば、IP アドレス、デバイス識別子、ログ タイムスタンプはデータ元のシステムによって異なる形式で提供されます。Zscaler Data Fabric は、これらを統一された形式に変換することで、パイプラインの後続のステップがスムーズに動作し、エラーや不一致を最小限に抑えます。

3. **強化：**形式化が行われると、データは追加のコンテキストで強化されます。これには、メタデータの統合、データ ポイントの関連付け、他のソースからのインサイトの追加が含まれます。たとえば、エンドポイントで検出された脆弱性は、資産の重要度や所有権、過去のリスク レベルに関する情報で強化される場合があります。このステップにより、データの品質が向上し、より有意義な情報が提供され、適切な意思決定をスピーディーに行えます。

4. **解決と重複排除：**Zscaler Data Fabric の主な強みの一つは、ソース間でデータを解決および重複排除できる点です。これは、エンティティー解決アルゴリズムを使用して、異なるログに表示される同じ IP アドレスなど、重複するレコードを特定してマージすることを目的としています。このステップは各エンティティーが一意に表現されるようにすることで、ノイズを減らし、データの精度を向上させ、セキュリティ イベントと資産の状態をより明確に理解できるようにします。

5. **グループ化：**最後のステップでは、関連するデータ ポイントをグループ化して、一貫性のあるエンティティーまたはクラスターを形成します。たとえば、複数の脆弱性アラートを資産に基づく単一のリスク プロファイルにグループ化できます。このアプローチにより、セキュリティ部門は関連するイベントやエンティティー、脅威を包括的に把握でき、より効率的な分析と迅速なインシデント対応が可能になります。

Zscalerのデータ ファブリック アーキテクチャーは、データファーストのアプローチを採用しており、最初に特定のユース ケースに焦点を当てることなく、データの集約、重複排除、正規化を行います。整理され、優先順位が付けられると、実用的なデータとなり、次のことが可能になります。

- **ビジネス ロジックの簡単な導入：**データ処理後に特定のビジネス ルールをデータ ファブリックに組み込み、プラットフォームが独自の運用ニーズに合致していることを確認できます。
- **ダッシュボードとレポートの迅速な構築：**データ ファブリックは、ダッシュボードとレポートの迅速な作成をサポートし、セキュリティの指標と傾向をリアルタイムで可視化します。
- **自動ワークフローへのシームレスな統合：**既存のワークフローに統合し、確立されたプロセスを中断することなく、データのインサイトに基づいて対応できるようにします。
- **リネージの維持：**データ ファブリックは、データ変換の完全な履歴を保持し、トレーサビリティーとコンプライアンスを確保します。
- **データを表示および変更できるユーザーの制御：**堅牢なロールベースのアクセス制御により、許可されたユーザーのみが機密データにアクセスし、変更できるようにします。
- **時系列での変化の把握：**Zscaler Data Fabric for Security はデータの変更を追跡するため、セキュリティ部門は傾向を分析し、異常を特定し、時間の経過に伴うデータの整合性を確保できます。

Zscaler Data Fabric for Security の中核となるデータ モデルの特長

Zscaler Data Fabric for Security は、柔軟性、適応性、精度を最適化するために設計された独自のデータ モデルを採用しています。ここでは、その特長を紹介します。

エンティティーベースのアーキテクチャー

従来のセキュリティ データ ソリューションは主にイベント ログに重点を置くのに対し、Zscaler のデータ モデルはエンティティーを中心に構築されています。エンティティーは、資産やユーザー、IP アドレス、組織のセキュリティ環境内にあるその他の個別の要素を表すものです。このアプローチにより、セキュリティ データについてより実用的な関係ベースのビューが提供されるため、より詳細なインサイトの入手とコンテキストを活用した分析が可能になります。Zscaler Data Fabric は、生のイベントだけでなくエンティティーに焦点を当てることで、関連付け、視覚化、リスク評価をより効果的に行います。

特定の原則と柔軟性の両立

Zscaler のデータ ファブリック モデルの主な原則の一つは、データの構造化、管理、アクセス方法について特定の原則と強固な仮定に基づいて構築されていることです。これは、規範的な構造を強調し、特定のユース ケースを優先する「構成よりも慣習」のアプローチを促進します。

カスタマイズ性と拡張性に優れたスキーマ

Zscaler のデータ ファブリック モデルは動的に設計されており、変化するセキュリティ ニーズに適応できるカスタマイズ可能なスキーマを備えています。組織は独自のエンティティー、属性、関係を定義できるため、新しいデータ ソースやセキュリティ ツールが統合されるたびにファブリックを調整できます。この拡張性は、データ モデルの関連性を維持し、進化するセキュリティ要件に合わせるうえで不可欠です。

経時的な進化

Zscaler Data Fabric for Security のセマンティック データ レイヤーは、集約された膨大な量のセキュリティ データを構造化および解釈するうえで重要な要素として機能します。セマンティック データ レイヤーにより、データの保存方法とアクセス方法を分離できるため、データ構造や要件の変化に合わせて進化できる柔軟なモデルが可能になります。このレイヤー内では、資産、チケット、結果、所有者などのエンティティー間の関係がマッピングされるため、バックエンド ストレージを理解しなくても、有意義にデータをクエリーできます。この抽象化は多様なデータ ソースを統合し、脆弱性管理やコンプライアンスなど、さまざまなセキュリティ アプリケーション全体で実用的なインサイトを生成するために重要です。Zscaler Data Fabric for Security は、データをセマンティック レイヤーに整理することで、セキュリティ情報の統一されたビューを作成し、大規模な再構成を行わずに新しいデータ ソースや進化するセキュリティ要件に迅速に適応できるようにします。

他のツールとのシームレスな統合

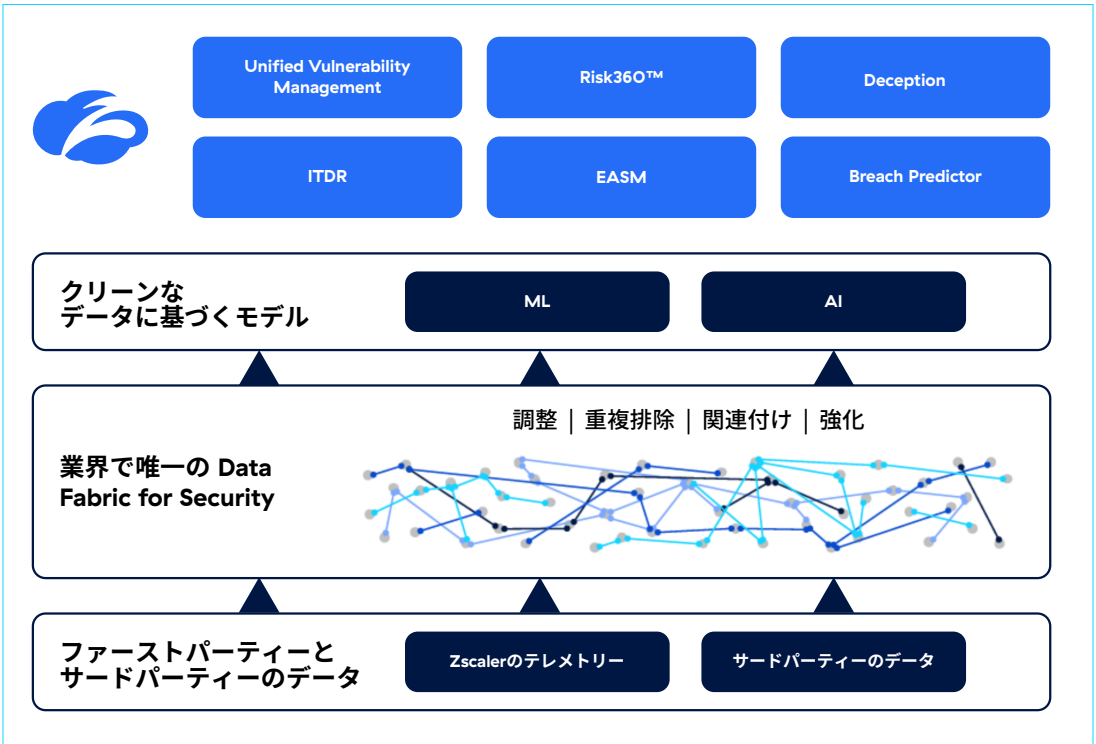
Zscaler Data Fabric のモデルは、SIEM、脆弱性管理システム、コンプライアンス ツールなど、さまざまなセキュリティ ツールとのシームレスな統合をサポートします。これは、API、コネクター、一般的なデータ形式の事前定義されたマッピングによって行われます。また、このモデルではエンティティーに基づいたインサイトをプラットフォーム間で簡単に共有できるため、サイロ化の解消と統一されたセキュリティ ビューの提供が可能になります。

迅速なアプリケーション開発のサポート

柔軟なデータ モデルにより、基礎となる同じデータ アーキテクチャーを活用してデータ ファブリック上に新しいアプリケーションを簡単に構築できます。たとえば、脆弱性管理、サイバー リスクの定量化、アイデンティティー管理のためのアプリケーションは、既存のデータ構造にすばやく接続できるため、開発時間が短縮され、カスタマイズ性と拡張性を兼ね備えたセキュリティ ソリューションをさらに迅速に展開できます。

Zscaler Data Fabric for Security を強化するソリューション

Zscaler Data Fabric for Security は現在、Unified Vulnerability Management (UVM) と Risk360 の基盤アーキテクチャーとして機能しています。今後、Zscaler Risk Management 全体でさらに他のソリューションもサポートしていく予定です。



ファブリック内で利用できる豊富な統合により、追加の Zscaler ソリューションがファブリックに接続され、製品間のフィードバック ループが提供されます。たとえば、Zscaler EASM は、外部の攻撃対象領域の調査結果と資産情報を UVM にフィードしてリスクの優先順位付けを行い、Risk360 にフィードしてサイバー リスクの定量化を行います。Zscaler DSPM は、リスク要因として使用するために UVM にデータ結果を提供し、UVM は修正が必要な要素に対して修復割り当てを送り返します。また、UVM や Risk360 がアクセス制御の変更を必要とするリスク インサイトを検知した場合、データ ファブリック内のコネクターを活用して、ポリシー調整の詳細を Zscaler ZIA および ZPA に送信します。

まとめ

今日の複雑なデータ環境においてデータ ファブリックの概念はいっそう重要になっており、多様なソースやプラットフォーム間でデータを統合、管理、活用する方法を提供しています。組織が生成するデータが指数関数的に増加し、リアルタイムの意思決定に対するニーズが高まるなか、シームレスな統合、ガバナンス、アクセシビリティーを実現するには、データ ウェアハウスやデータ レイクなどの従来のデータ ソリューションでは不十分です。データ ファブリックのメタデータによるリアルタイムの統合は、データ アクセス、ガバナンス、分析を可能にする単一の一貫したレイヤーを提供することでこれらの課題に対処するため、戦略的な意思決定に不可欠な要素となっています。

サイバーセキュリティの分野では、データの断片化が依然として最も重要な課題の一つです。さまざまなツールが生成する大量のデータは統合されておらず、結果的に可視性が欠如し、対応に遅れが生じます。これは多くのセキュリティ部門が直面している問題です。Zscaler Data Fabric はセキュリティ エコシステム全体のデータを集約するだけでなく、強化し、関連付ける統合レイヤーとして機能することで、この問題を解消します。Zscaler Data Fabric が構築する「信頼できる唯一の情報源」により、迅速な脅威検出、リスクの優先順位付け、新たな脅威へのリアルタイムの対応が可能になります。

Zscaler のアプローチは従来の SIEM やその他のセキュリティ ツールとは大きく異なり、エンティティーベースのアーキテクチャーを基盤としたセキュリティ重視のデータ ファブリックを構築します。そして、データの品質、コンテキスト、相互運用性を優先し、より深いインサイトとより正確なリスク評価を可能にすることで、さらに適応性が高く、効率的でスケーラブルなセキュリティ戦略を実現します。

最終的に、組織がデータ管理とセキュリティ運用の最適化を目指すなかで、Zscaler のようなデータ ファブリックの採用は、有益であるだけでなく、必要になります。データ ファブリックはデータ サイロを解消するだけでなく、分析機能を強化し、リアルタイムの意思決定を推進することで、データ中心の考え方が浸透した現代において組織をより効果的に保護できるようにセキュリティ部門をサポートします。

用語

- **データ ファブリック：**多様なシステムや環境間でシームレスなデータ統合、管理、アクセスを可能にするアーキテクチャーで、データの統一されたビューを提供して、分析、ガバナンス、意思決定を改善します。
- **Data Fabric for Security:** セキュリティのユース ケースへの対応に関連するエンティティーをサポートするために特別に設計されたデータ ファブリックです。
- **エンティティー：**システムやプロセス、データベース内で識別可能な属性を持つ個別のユニットを指します。エンティティーは、システム設計、データベース、セキュリティ フレームワークの基本的な要素であり、管理、監視、または保護する必要があるさまざまな要素を定義します。エンティティーは、サーバー、ルーター、プログラム、プラットフォーム、ユーザー、所有者など、さまざまな要素を表すことができます。
- **エンティティー解決：**属性のバリエーションに関係なく、同じエンティティーを表すデータベース間でレコードを識別してマージするプロセスです。
- **アウトグレーション：**他のシステムへのアウトバウンド コネクターまたは統合を指す Zscaler の用語で、通常はチケット発行システムやプロジェクト管理システムを指しますが、データ レイクや SIEM、組織が運用データを送信する必要があるものすべてを指すこともあります。
- **セマンティック データ レイヤー：**共通の用語と関係を定義することで、データをビジネスに適した用語や概念に変換して統合ビューを提供する抽象化レイヤーであり、アプリケーション間でのデータの理解と使用を促進します。
- **ETL (抽出、変換、ロード):** ソースからデータを抽出し、目的の形式に変換して、分析のために対象のシステムまたはデータベースにロードするデータ統合プロセスです。
- **Open Security Graph™:** Zscaler Data Fabric for Security およびその拡張性とカスタマイズ性に優れたデータモデルを活用して、さまざまなエンティティー間の関係を包括的かつ視覚的に表現します。

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