



データ統合の 将来像

すべての必要なデータに簡単にアクセスし、
有効に活用するには



目次

はじめに.....	3
第 1 章：手動によるデータ統合の課題	6
第 2 章：データ統合の障壁を打ち破る	8
第 3 章：AWS でデータ統合が簡単に	10
1. サービスを統合してゼロ ETL な未来を実現する.....	11
2. 付加価値をつけるデータ変換とデータパイプラインを AWS Glue で簡単に実行する	18
3. 無数のデータソースに接続する	21
4. データ共有を安全かつ簡単に行う	23
まとめ：AWS でのデータ統合でデータの価値を引き出す	24



はじめに

さまざまなツールやシステムで保存、分析されたデータを統合できれば、顧客やビジネスをより深く理解できます。何が起こったのか、なぜ起こったのかを知ることができ、次に何が起こるかを予測できるようになり、その確信も強くなっていきます。劇的な革新をもたらすデータドリブンなインサイトは、これまでの状況と今後の方向性の全容を把握することで得られます。

例えば、あるホテルチェーンのマーケティング部門を運営しているとしましょう。ハイバリューカスタマーの宿泊体験を向上させるため、ターゲットを絞ったオファーの作成を検討しているとします。リレーショナルデータベースには顧客の購入履歴、分析システムにはホテルのウェブサイトからのクリックストリームデータ、サポートシステムには顧客チャットの記録があります。こうしたデータセットを利用して、顧客が競合ホテルに予約する可能性が高い時期を予測する機械学習モデルを構築し、適切なインセンティブを適切なタイミングで提供できるようにしたいと考えています。

この例からわかるのは、完全な顧客プロフィールを利用してタイムリーな予測を行うには、3つのデータセットすべてを統合する必要があるということです。全体像を把握し、さまざまなデータを真のビジネス価値へと変換するには、データ統合が鍵となります。しかし、色々なツールに保存された異なるソースからのデータをまとめるのは困難です。組織がサイロを使用しており、それによってデータにアクセスしにくかったり、システム間に距離が生まれたり、各レベルでユーザーがデータへのアクセスを妨げられている場合は、さらに困難になります。





長い間、データ統合は手間のかかる作業であり、生産性の低下、コストの上昇、継続的なエラーの一因となっていました。多くのデータ部門は、別々のデータサイロにあるデータを統合するために、複雑な抽出、変換、ロード (ETL) パイプラインを構築しなければならず、完成までに数日とは言わないまでも数時間はかかります。そしてこれは始まりに過ぎません。いったん ETL パイプラインを構築すると、それを維持するためにさらに多くの時間と労力を費やす必要があります。パイプラインを継続的に管理して、データが最新かつ適切なものであることを確認する必要があります。また、パイプラインを運用化し、ダウンタイムを回避するために集中的に取り組む必要もあります。これはよく、スケジューリング、モニタリング、トラブルシューティングの終わりのないループとして現れます。

ETL は現状の維持はできるかもしれませんが、意思決定を行うスピードとしては不十分です。ETL は、よりシンプルにする必要があります、多くの場合は廃止する必要があります。

Amazon Web Services (AWS) の目標は、データがどこにあっても、すぐにすべてのデータを接続して活用できるようにすること、そしてそれをデータドリブンな意思決定に必要なスピードと自信を持って簡単に行えるようにすることです。AWS では、効果的なデータ統合として 4 つの項目を据えています。第一に、AWS のサービス間を直接統合し、一般的なユースケースでの ETL を削減および排除しています。これがチームの俊敏性向上につながります。その一つとして、AWS はデータの移動、ロード、前処理を行うデータパイプラインの構築や管理をせずに、分析、機械学習、ビジネスインテリジェンス (BI) を実行できる、**ゼロ ETL** な未来に向けて投資しています。

第二に、複数のタイプのデータセットを組み合わせたり、変換で付加価値をつけたりなどのユースケースで ETL が必要な場合は、**AWS Glue** を使用して ETL を簡単に行えるようにしています。

第三に、一部のデータだけでなく、すべてのデータを活用できるように、AWS のサービスは、サードパーティーの Software as a Service (SaaS)、オンプレミス、その他のクラウドなど、数百ものデータソース (その数は増え続けています) に接続してフェデレートでき、さらにサードパーティーデータともシームレスに統合します。

第四に、データで何かを行いたい場合は、AWS でデータを共有して読み取り/書き込みアクセスが可能になるため、複数のチームが同じ場所にあるデータを活用できます。ワークロードを分離して分析のパフォーマンスを向上させたり、データを移動またはコピーすることなくチーム間のコラボレーションを促進したりできます。

この日本語ガイドでは、ETL パイプラインの手動での構築と維持管理に関連する一般的な課題を詳しく見ていきます。そして、これらの課題を適切なデータ統合テクノロジーによってどのように解決できるのかを解説します。そのうえでまとめとして、データ統合の 4 つの項目について詳しく説明します。

AWS がデータ統合を迅速かつ簡単にしている 4 つの方法

1. AWS サービス間の直接統合によって、一般的なユースケースでの ETL を削減および排除
2. 付加価値をつけるためのデータ変換などでの ETL を AWS Glue を使用して簡易化
3. 無数のデータソースやサービスとの接続およびフェデレーションによりパートナーやサードパーティーのデータを扱う
4. 安全なデータ共有を可能にしてコラボレーションを強化



第 1 章

手動によるデータ統合の課題

従来の ETL プロセスは、障害物コースのようなものと言えます。例えば、複数の国に数十の工場を持つグローバルな製造会社を考えてみましょう。データベースのクラスターを使用して、各国の注文および在庫データを保存しています。注文と在庫をリアルタイムで把握するには、これらの各データベースクラスターと中央データウェアハウス間に個別のデータパイプラインを構築して、結合されたデータセット全体にクエリを実行する必要があります。これを実現するために、データ統合チームは異なる 12 のクラスターに接続するコードを作成し、12 の稼働パイプラインを管理およびテストしなくてはなりません。デプロイされた後は、パフォーマンスを最適化するためにパイプラインを常にモニタリングしてスケールする必要があります。何か変更があれば、12 か所すべてでアップデートが必要です。

ETL は複雑で時間がかかり、コストがかさむものとして広く認知され、敬遠されています。目的のターゲットスキーマに一致するようにデータをマッピングするには、複雑なデータマッピングルールが必要であり、データの不整合や競合の処理をしなくてはなりません。エンジニアは問題を診断するために、効果的なエラー処理、ログ記録、通知のメカニズムを実装する必要があります。データセキュリティ要件によって、さらに厳しいシステムの制約も生じます。





こうしたことを実現するには、ETL パイプラインの構築と維持管理の専門スキルを持ったエンジニアチームが必要です。パイプラインを構築するためにカスタムコードを作成するデータエンジニアと、パイプラインをスケールするためのインフラストラクチャをデプロイおよび管理する DevOps エンジニアが求められます。このチームがパイプラインの構築を完了するには、数日ではないにしても何時間もかかります。そしてデータソースが変更されるたびに、プロセス全体を繰り返す必要があります。

また、構築フェーズの間はデータが利用できないため、データアナリスト、データサイエンティスト、その他のエンドユーザーは作業を中断することになります。組織全体でリアルタイムの意思決定ができなくなり、ほぼリアルタイムで対応すべきオンライン広告の掲載、不正トランザクションの検出、リアルタイムのサプライチェーン分析といったユースケースにデータが対応できなくなる可能性があります。

また、最初の ETL 構築も面倒ですが、それで労力と費用がなくなることはありません。実際、ETL の費用はデータ量が増えるほど膨らみます。システム間でデータストレージが重複している場合、大量のデータを保存するために、無駄な費用が発生する可能性があります。さらに、ETL プロセスのスケールには、多くの場合、コストのかかるインフラストラクチャのアップグレード、クエリパフォーマンスの最適化、並列処理技術が必要になります。要件が変更された場合、データエンジニアリングチームはアップデート処理中のパイプラインを継続的にモニタリングおよびテストする必要があり、メンテナンスコストが増加します。

第 2 章

データ統合の障壁を打ち破る

データソースはパズルのピースのようなものです。データ統合は、断片化されたこれらのピースを集めて継ぎ目なくまとめ、データを単一で統一されたビューとして提示します。このビューにより、組織は顧客とビジネスについてより深く理解できます。ただし、従来の ETL プロセスでは、このような全体像をある程度迅速に、または確度を持って明らかにすることが困難です。

AWS では、ETL パイプラインの構築と管理における画一的な作業を自動化し、すべてのデータをよりスピーディーに統合して活用できるように取り組んでいます。AWS のデータ統合テクノロジーは、データパイプラインの構築に費やす時間とリソースを削減し、ユーザーがより迅速にデータにアクセスできるようにします。AWS のサービスは、データアーキテクチャを簡素化し、データエンジニアリングの労力を軽減します。継続的なコストと反復的な作業に捕われることなく、生産性を向上し、価値の高いクリエイティブな仕事に集中できるようにします。

例えば、AWS のゼロ ETL 統合は[クラウドネイティブでスケーラブル](#)であるため、実際の使用量とデータ処理ニーズに基づいてコストを最適化できます。インフラストラクチャのコスト、開発作業、維持管理のオーバーヘッドを削減できます。また、ゼロ ETL では、大量のデータを再処理しなくても新しいデータソースを含めることができ、反復的な作業が不要になります。

ゼロ ETL を使用すると、ソースからターゲットへのデータ移動も自動化され、労力もかかりません。データにほぼリアルタイムでアクセスできるようになり、最新のデータを分析、AI、機械学習、レポートに確実に投入できます。ビジネスインサイトをより迅速に発見し、重要な瞬間に判断を下すことができます。この即時性は、ニアリアルタイムのダッシュボード、データ品質モニタリング、顧客行動分析などのユースケースにおいて大きな意味があります。

データの統合はイノベーションにとって欠かせませんが、重要なことは、これが技術的および運用上の利点だけを目的とするものではないということです。データ統合は、文化的にも意義があるものです。ほとんどのデータリーダーにとって、データドリブンなカルチャーを確立することが至上の目標です。組織内のチームがデータを信頼し、リアルタイムで使用してユーザーエクスペリエンスを変革すると、そのようなカルチャーの形成と強化が自然に始まります。

AWS のデータ統合テクノロジーがどのようにイノベーションをスピードアップするのか

- データパイプラインの構築に費やす時間とリソースを削減する
- より早くデータにアクセスできるようにする

ゼロ ETL 統合を利用したお客様の導入事例

KINTO

KINTO テクノロジーズ株式会社は、モビリティプラットフォーム業界のリーディングカンパニーであり、トヨタの金融サービス企業として KINTO サービスの開発を担当するテクノロジー企業です。Amazon Aurora MySQL zero-ETL Integration with Amazon Redshift を使用し、同社はより回復力のあるデータパイプラインを実現しました。Amazon Redshift の高度な分析機能を、ほぼリアルタイムで運用データに適用できるようになりました。

「Zero-ETL 統合が利用可能になる前は、コアデータベースからダウンロードのアプリケーションに変更を継続的にストリーミングするカスタムビルドのソリューションを使用していましたが、パフォーマンス上の課題と本番ワークロードへの影響に常に直面していました。本番環境のワークロードへのパフォーマンス影響に対処するため、更新の送信頻度を減らすようパイプラインを調整する必要があり、直近ではない古いデータを使用しなければならませんでした。Aurora MySQL zero-ETL Integration with Amazon Redshift を使用することで、Amazon Redshift でいつでもほぼリアルタイムのデータを利用できるようになり、開発者が手動で ETL 運用のデータパイプラインを管理したり、ワークロードのパフォーマンスへの影響に対処したりする必要がなくなり、運用上の負担を軽減できます」

KINTO テクノロジーズ株式会社、
取締役副社長、
景山均 氏



第 3 章

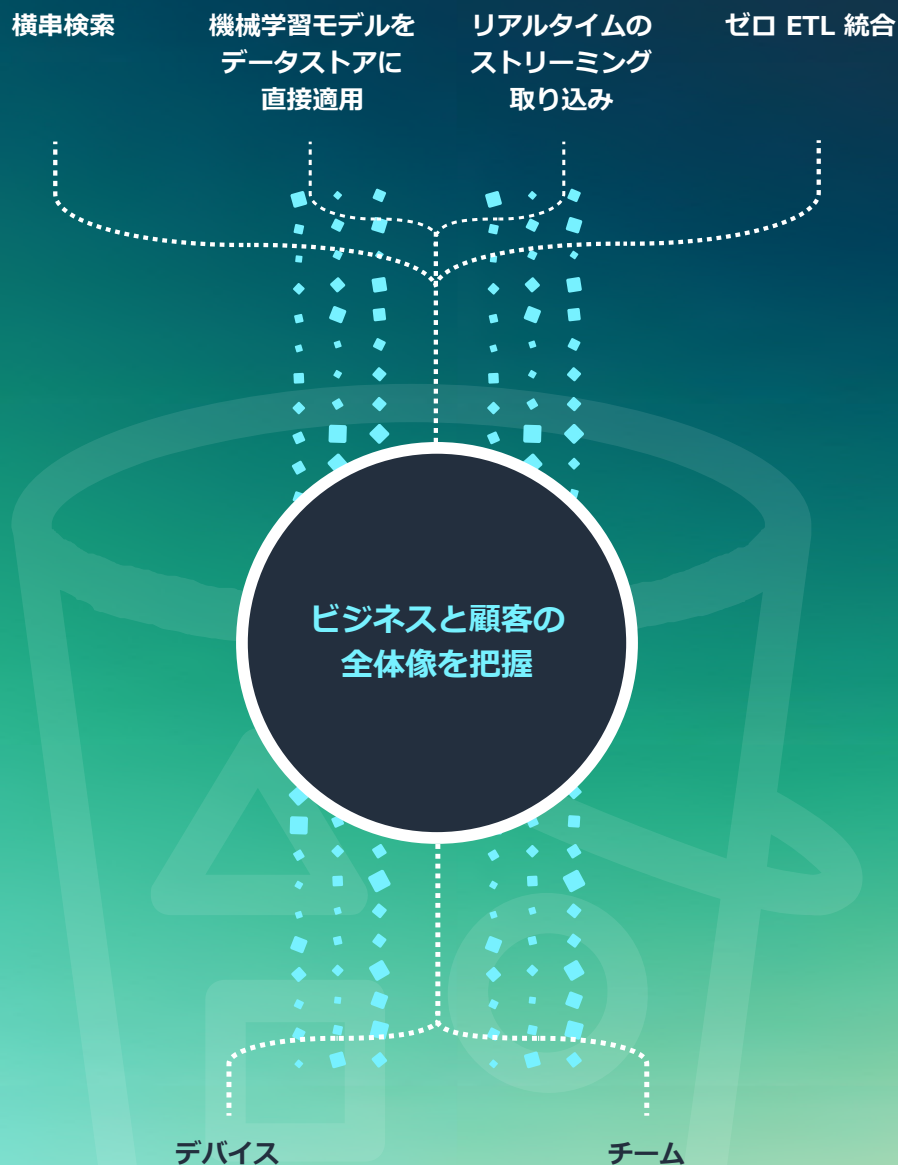
AWS で データ統合が 簡単に

AWS は、どんなデータも保存場所に関わらず、すばやく簡単に統合して活用できる未来に投資しています。この日本語ガイドの冒頭で概説したように、AWS のデータ統合アプローチには 4 つの柱があり、次のことを簡単に実現できるようにします。

- 1 サービスを統合してゼロ ETL な未来を実現する
- 2 付加価値をつけるデータ変換とデータパイプラインを AWS Glue で実行する
- 3 無数のデータソースに接続する
- 4 データ共有を安全かつ簡単に行う



手動パイプラインの必要性を排除



1. サービスを統合してゼロ ETL な未来を実現する

ゼロ ETL な未来では、データの移動、ロード、前処理を行うデータパイプラインを手動で構築または管理しなくても、分析、機械学習、ビジネスインテリジェンス (BI) を実行できるようになります。AWS は、手動のデータパイプラインが不要の数々のユースケースを発表し、この未来を実現しています。

横串検索

[Amazon Redshift](#) と [Amazon Athena](#) の横串検索を使用すると、運用データベース、データウェアハウス、データレイクに保存されているデータを移動せずに予測分析が実行できます。

横串検索を使用すると、使い慣れた SQL コマンドを使用して複数のデータソースのデータを結合してすばやく分析し、結果を [Amazon Simple Storage Service \(Amazon S3\)](#) に保存して後で使用できます。こうして複雑な ETL パイプラインを使わずに、データを柔軟に取り込みます。

データストアで直接使用できる機械学習モデル

[Amazon SageMaker](#) は AWS のデータウェアハウスやデータベースと統合されます。そのため、データパイプラインの構築は必要なく、機械学習の専門知識がなくてもデータを機械学習に利用できます。データによるモデルのトレーニングや、推論結果のデータストアからの直接追加などに利用できる、さまざまな統合を選択できます。これらはデータのエクスポートや処理をしなくても実行できます。

リアルタイムのストリーミング取り込み

AWS ストリーミングサービスを直接統合することで、データが生成されたらすぐに分析し、インサイトをタイムリーに収集してチャンスを逃さず活かすことができます。例えば、[Amazon Redshift ストリーミングの取り込み](#)では、Amazon Redshift の設定によって、Amazon Redshift コンソールからストリーミングデータをリアルタイムでデータウェアハウスに直接取り込みます。この統合機能では、1 秒間に数百メガバイトのデータを取り込み、ニアリアルタイムでデータをクエリできます。また、データを Amazon S3 にステージングしなくても、[Amazon Kinesis](#) のデータストリームや [Amazon Managed Streaming for Apache Kafka \(Amazon MSK\)](#) のデータストリームに接続して、データを Amazon Redshift に直接取り込むこともできます。

ゼロ ETL 統合

AWS では、AWS の最も人気のあるデータストア (Amazon Redshift と統合された 4 サービス、[Amazon OpenSearch Service](#) と統合された 2 サービスほか) で、一般的な ETL ジョブ向けのゼロ ETL 統合を用意しています。

これらのゼロ ETL 統合を利用すると、データはソースから送信先に自動的に接続されるため、トランザクションデータをすばやく分析できます。また、パイプラインを開発する必要がないため、必要なインサイトを得るためにパイプラインが構築されるのを待つ必要はありません。データエンジニアリングチームの数か月分の作業が不要になり、より付加価値の高い活動に集中できるようになります。同時に、データドリブンな予測をより迅速かつ正確に実行でき、コンテンツターゲティング、不正検出、顧客行動分析などに利用できます。

これらの統合は使いやすく、セットアップもシンプルです。ソースを選択し、ターゲットを選択するだけです。また、複数のソースからのデータをシームレスに統合できるため、複数のアプリケーションやデータソースにわたって統合分析や検索を実行できます。

Amazon Redshift および [Amazon OpenSearch Service](#) それぞれのゼロ ETL 統合の概要と、その使用方法を次に挙げます。

AWS のゼロ ETL 統合の利点

- インサイトへの迅速なアクセスを実現
- データエンジニアリングチームの数か月分の作業を排除
- 使用が簡単
- 複数のソースからのデータを統合

AMAZON REDSHIFT とのゼロ ETL 統合

ゼロ ETL 統合	説明	特長	ユースケース
Amazon Aurora MySQL 互換および PostgreSQL 互換	<u>Amazon Redshift との Amazon Aurora ゼロ ETL 統合</u> では、Amazon Redshift を使用してニアリアルタイム分析や機械学習が可能で、Aurora からのペタバイト規模のトランザクションデータを分析できます。	- 複数の Aurora クラスターから 1 つの Amazon Redshift ウェアハウスにデータをレプリケートする - 本番環境のワークロードに影響を与えることなく、アプリケーション全体の包括的なインサイトを取得する - ソースの Aurora クラスターで行われたスキーマの変更を Amazon Redshift に自動的に反映するアプローチであるため、ETL よりも適応力があり安定性が非常に高い	コンテンツターゲットティング
Amazon RDS for MySQL	<u>Amazon Relational Database Service (Amazon RDS) for MySQL と Amazon Redshift の統合</u> により、RDS for MySQL データを簡単に分析できます。	- RDS for MySQL データを Amazon Redshift にシームレスに複製し、初期データロード、進行中の変更の同期、スキーマのレプリケーションを自動的に処理する - ワークロードを分離して最適なパフォーマンスを実現する - Aurora MySQL 互換エディションや Aurora PostgreSQL 互換エディションなど、複数のソースからのデータを Amazon Redshift に統合する	- 最適化されたゲーム体験 - データ品質のモニタリング - 不正行為の検出
Amazon DynamoDB	<u>Amazon Redshift との Amazon DynamoDB ゼロ ETL 統合</u> により、 <u>DynamoDB からのデータを Amazon Redshift で分析に使用するフルマネージド型のソリューションが可能になります。</u>	- DynamoDB 読み取り容量ユニット (RCU) を消費することなく、DynamoDB データを Amazon Redshift に複製して分析する - 本番環境のワークロードに影響を与えることなく、アプリケーション全体の包括的なインサイトを取得する - 高速 SQL クエリ、機械学習統合、高速集計のためのマテリアライズドビュー、安全なデータ共有など、強力な Amazon Redshift 機能を DynamoDB データに活用できる	顧客行動分析

AMAZON OPENSEARCH SERVICE とのゼロ ETL 統合

ゼロ ETL 統合	説明	特長	ユースケース
Amazon DynamoDB	Amazon OpenSearch Service との <u>Amazon DynamoDB</u> ゼロ ETL 統合により、DynamoDB に保存されたデータに対するウェブサイト検索や製品検索など、アプリケーション検索が構築できます。	- 強力なフルテキストクエリやベクター検索クエリを DynamoDB データにニアリアルタイムで簡単に実行する - DynamoDB に書き込まれるデータを数秒以内に OpenSearch Service に複製する - 複数の DynamoDB テーブルのデータを 1 つの OpenSearch Service マネージドクラスターまたはサーバーレスコレクションに同期して、複数のアプリケーションにわたる包括的なインサイトを取得し、検索アセットを統合する	質の高い検索エクスペリエンスを創出する
Amazon S3	Amazon OpenSearch Service との <u>Amazon S3</u> ゼロ ETL 統合により、Amazon S3 に保存されているクエリ頻度の低いログデータを分析して、すべてのデータに対してセキュリティ分析や運用分析を実行できます。	- 運用データ分析ツールを切り替えることなく、Amazon S3 および Amazon S3 ベースのデータレイクに保存された運用ログをクエリする新しい手法を提供する - クエリのパフォーマンスを強化する - 組み込みのクエリアクセラレーション機能を使用して高速でロードするダッシュボードを構築できる - データを移動せずに、複雑なクエリと視覚化を実行する	- セキュリティとログデータを分析する - 機密データを保護する

ゼロ ETL 統合を利用したお客様の導入事例



Money Forward

Money Forward は直感的で使いやすい SaaS 管理プラットフォームである Admina を IT 部門向けに提供しています。Admina は IT 部門での反復作業の合理化、コスト削減、セキュリティ強化に役立ちます。同社は、Amazon Aurora MySQL からの製品データを Amazon Redshift で分析するための ETL オペレーションを Admina に実装しメンテナンスすることに苦慮していました。Aurora MySQL を Amazon Redshift とゼロ ETL 統合することで、Aurora MySQL データベースと Amazon Redshift 間のデータ同期をニアリアルタイムで行えるようになり、1 か月かかっていた分析環境の構築をわずか 3 時間に短縮できました。ゼロ ETL 統合により、開発時の初期負担が軽減されただけでなく、本番環境への影響も少なくなりました。

「Amazon Aurora と Amazon Redshift のゼロ ETL 統合がリリースされる前は、Amazon Redshift で Amazon Aurora MySQL の製品データを分析するための ETL オペレーションの実装とメンテナンスの負担は大変でした。Aurora MySQL を Amazon Redshift とゼロ ETL 統合することで、当社の Aurora MySQL データベースと Amazon Redshift 間でほぼリアルタイムのデータ同期が可能になり、分析環境の構築にかかる時間が 1 か月からわずか 3 時間に短縮されました。開発時の初期負担が軽減されるだけでなく、ゼロ ETL 統合により本番環境への影響が少なく、最小限のコストと最大速度で分析環境を構築できます」

マネーフォワード i 株式会社
取締役兼最高執行責任者
村上勝俊 氏





ゼロ ETL 統合を利用したお客様の導入事例

W WOOLWORTHS

Woolworths は、イノベーション、価値、サステナビリティに重点を置き、高品質の衣料品、雑貨、食品を幅広く提供するサブサハラアフリカの大手小売業者です。ビジネス全体でデータドリブンな意思決定を促進し、重要で時間的制約があるイベントに効果的に対応するには、データからタイムリーなインサイトを引き出すことが重要です。Amazon Aurora MySQL を Amazon Redshift とゼロ ETL 統合することで、同社は開発期間を 2 か月から 1 日に短縮しました。ゼロ ETL 統合によって、データがいつでもクエリできるようになり、データレイテンシーを大幅に減少できました。こうして同社は、イベント発生時の意思決定をより速やかに行えるようになりました。また、エンジニアリングの労力を軽減し、パイプライン管理における障害点を減らし、コストを削減しました。

ゼロ ETL 統合を利用したお客様の導入事例

INTUIT

Intuit は、TurboTax、Credit Karma、QuickBooks、Mailchimp を通じて、1 億にのぼる利用者および中小企業の顧客の成功を支援するグローバルな金融テクノロジー企業です。同社は、1 日あたり 1,000 万件超のプロファイル移行という非常に大きなタスクを含んだ移行計画の準備を進めていました。データインGEST処理を合理化し、複雑なエンジニアリング作業をなくすために、Amazon Aurora MySQL のゼロ ETL 統合を利用しました。これによって、データのキャプチャや個別の取り込みジョブを必要とせずに、大量のデータを Amazon Redshift に送信できました。また、重要な技術的およびビジネス上の意思決定に役立つ迅速なインサイトが得られ、数か月分の作業を削減できました。同社は、大規模なデータ移行やニアリアルタイム分析の新しいパターンを検討できるようになりました。



2. 付加価値をつけるデータ変換とデータパイプラインを AWS Glue で簡単に実行する

特定のユースケースでは、ETL パイプラインの構築が引き続き必要です。データ分析や機械学習モデルの作成のために、データエンジニアがデータクレンジングや重複排除などのデータ変換を実行したり、カスタムアプリケーション間で複数のデータセットを組み合わせたりする必要が出てくることもあるでしょう。こうしたユースケースのデータ変換を容易にするのが、サーバーレスでスケーラブルなデータ移動と変換を行うサービスである [AWS Glue](#) です。

AWS Glue は、データとデータパイプラインを接続、変換、管理するフルマネージド型のデータ統合サービスです。毎月、数十万のお客様によって利用されており、何億ものデータ統合ジョブが実行されています。AWS Glue を使用すると、データ統合プロセスの簡素化により、データがすぐに利用でき、さまざまな分析アプリケーション向けに適切にフォーマットされるようになります。

また、スケーラビリティも AWS Glue の要です。ETL ジョブの量と複雑さに合わせて AWS Glue がリソースを自動で割り当てるため、お客様はインフラストラクチャの管理をすることなく、ペタバイト規模のデータからのインサイトの取得に集中できます。こうした適応型スケーリングを備えた AWS Glue は、組織の規模に関わらず利用できる効率的なソリューションであり、さまざまなデータソースとフォーマットをシームレスに統合して包括的な分析を可能にします。その結果、情報に基づいたインテリジェントなビジネス上の意思決定をサポートするのに不可欠な、データ統合プロセスを合理化できます。

AWS Glue では、生成 AI を活用してデータをより迅速に統合することもできます。[Amazon CodeWhisperer](#) との統合により、[AWS Glue Studio](#) でコードの提案や構文の修正がリアルタイムで提示されるので、経験のあるビルダーはノートブックを開いたまま、ベストプラクティスやコード提案を参照できます。AWS Glue に [Amazon Q](#) のデータを統合すると、自然言語を使用してデータ統合パイプラインを構築できます。

あらゆるデータを大規模に検出、準備、統合する



オールインワンの
データ統合
サービス



すべてのデータ
ユーザーをサポート
するカスタム
ツール



コスト効率に優れ、
サーバーレスで
スケーラブル



すべての
ワークロードを
1 か所でサポート



生成 AI 機能で
データをより
迅速に統合



AWS GLUE を利用したお客様の導入事例



BMW は、多くの自動車会社と同様に、世界的な半導体不足によりサプライチェーンの課題に直面しています。サプライヤーや製造メーカーにおける半導体不足を解決するには、同社の現在および今後の半導体需要の見通しを立てることが重要です。同社は AWS Glue、Amazon S3、その他の AWS サービスを使用して、確かな見通しを持った、長期的な半導体需要の自動予測を実現しました。AWS Glue を使用して多数のデータソースからデータを取り込み、マスターテーブルに集約してデータをクレンジングし、サードパーティー製のシステムを使用するデータコンシューマーが利用できるようにしました。

AWS GLUE を利用したお客様の導入事例



Itaú Unibanco は、ラテンアメリカの最大手銀行の一つであり、6,500 万以上の顧客にサービスを提供しています。投資プラットフォーム、ホールセール業務、保険、当座預金口座など、複数のサービスを提供しています。通常、同行の全事業部門の支払いシステムを把握する包括的なビューの生成には数か月かかります。同行は AWS Glue などの AWS サービスを活用したデータメッシュアーキテクチャを実装し、顧客支払いの全データを 24 時間以内に処理、分析できるようになりました。

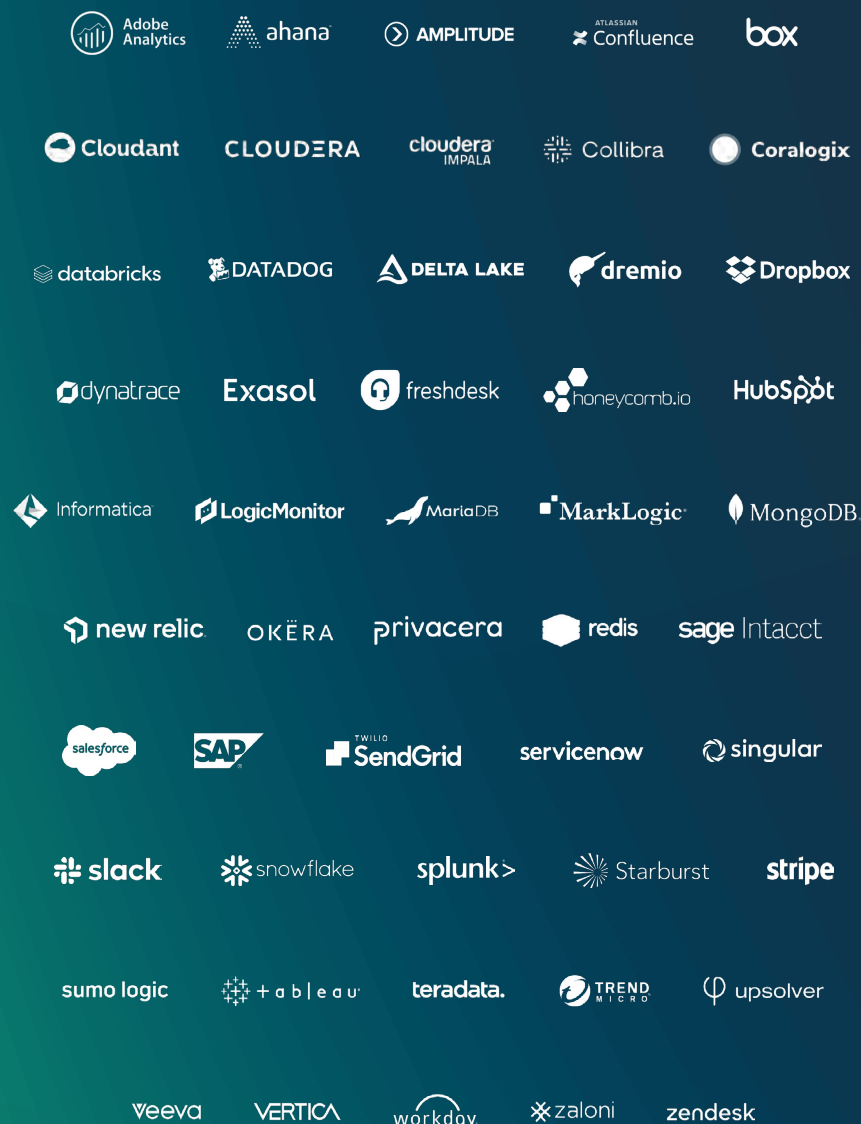


3. 無数のデータソースに接続する

組織が一部のデータだけでなく、すべてのデータを活用できるように、AWS のサービスは、サードパーティーの SaaS、オンプレミス、その他のクラウドなど、数百ものデータソース (その数は増え続けています) に接続し、サードパーティーのデータともシームレスに統合しています。AWS では、SAP などの ERP アプリケーションから Salesforce などの CRM アプリケーション、Adobe Analytics などの分析サービスまで、社内全体にまたがるデータソースに接続できます。

このような接続を可能にする AWS のサービスの一部をご紹介します。

- [Amazon AppFlow](#): データレイクとデータウェアハウスを 50 種類以上の SaaS アプリケーションに接続します
- [Amazon Kinesis Data Firehouse](#): 30 を超える AWS およびサードパーティーのソースから、データをリアルタイムでストリーミングします
- [Amazon Athena](#): 25 を超えるデータソースにクエリを実行します
- [Amazon SageMaker Data Wrangler](#): 40 を超えるソースのデータにアクセスして、機械学習モデルを構築できます
- [Amazon QuickSight](#): 30 を超えるソースを使用してインタラクティブなダッシュボードを構築できます
- [AWS DataSync](#): データの AWS 内への移動および AWS 外への移動を高速化し、ハイブリッド環境での処理を支援します
- [AWS Glue](#): 無数のデータソースからデータを取り込みます
- [Amazon Managed Workflows for Apache Airflow \(Amazon MWAA\)](#): コミュニティが提供する多数の Airflow オペレーターやセンサーからのデータパイプラインを定義します



サードパーティーデータについては [AWS Data Exchange](#) を提供しており、これ 1 つで 300 を超えるデータプロバイダーと 3,500 を超えるデータ製品のデータのすべてをファイル、テーブル、API 経由で利用できます。AWS のデータ、分析、機械学習サービスとすばやく統合でき、クラウドですぐに使えるデータを簡単に見つけてサブスクライブできます。

さまざまな業界のお客様が AWS Data Exchange のサードパーティーデータを使用しています。組み合わせの例としては、平均寿命ベンチマークデータを使用して新薬を研究する製薬会社、位置データをサブスクライブして事業拡大すべき場所を特定するレストラン、気象データを活用して顧客のニーズを予測し在庫を最適化する小売業者などがあります。

AWS Data Exchange は、AWS にネイティブに統合されるため、データを簡単に使用できます。例えばサードパーティーのデータファイルを Amazon S3 にダイレクトに取り込めます。また、Amazon Redshift テーブルによるデータ配信も依頼でき、プロバイダーにデータのクレンジング、検証、実稼働用テーブルへの変換に必要な作業を行ってもらうことで、お客様はサブスクライブ後すぐにデータをクエリ、分析、稼働システムとの統合に利用できます。API によるデータ配信も依頼でき、デベロッパーはそのデータを本番アプリケーションが構築されている場所にそのまま統合できます。

アプリケーション、分析、機械学習 モデルでサードパーティーデータを すばやく簡単に使用する



幅広いデータ
セットの選択肢



優れたデータ
テクノロジー



合理化された
データの調達と
データガバナンス

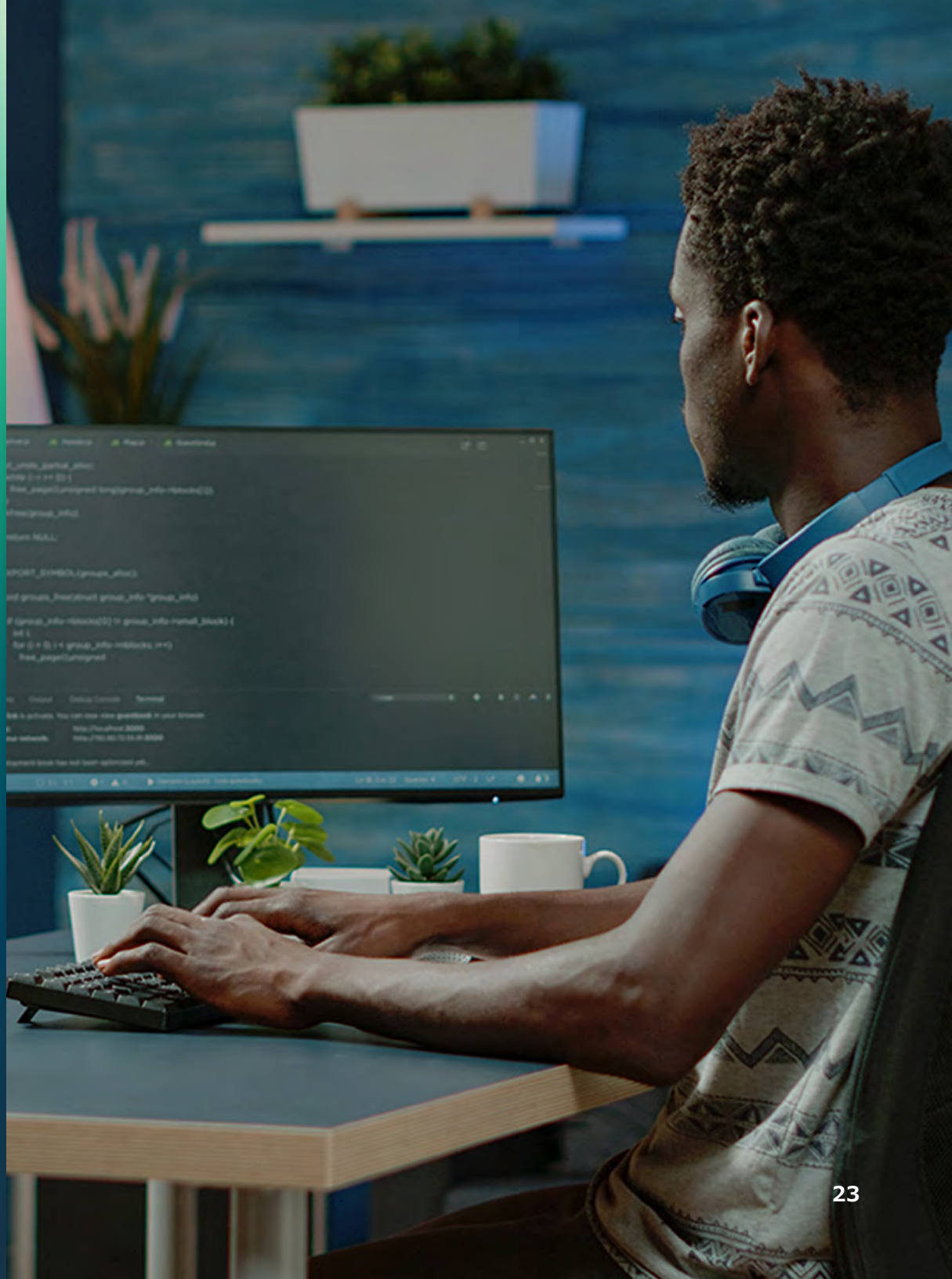


使いやすい
データ分析と
機械学習

4. データ共有を安全かつ簡単に行う

データをパートナーと共有するには、安全で効果的な方法が必要です。[AWS Clean Rooms](#) では、お客様とそのパートナーが、双方の基データの共有やコピーを行うことなく、また機密情報を互いに開示することなく、集合データセットを使用して、簡単かつ安全に共同作業や分析、機械学習モデルの構築ができます。安全なデータルームは数分で作成できます。AWS クラウドで他の企業とコラボレートして、広告キャンペーン、投資判断、研究開発に関するユニークなインサイトを得られます。

また、[AWS B2B Data Interchange](#) を利用すれば、電子データ交換ドキュメントを一般的なデータ形式に自動的に変換できるため、B2B トランザクションデータを準備しビジネスアプリケーションに統合する煩雑さと、作業にかかる費用を減らせます。同様に、Amazon Redshift のデータ共有では、データを移動またはコピーすることなく、組織内、組織間、AWS リージョン間、さらにはサードパーティープロバイダー間でデータを共有できます。



まとめ

AWS でのデータ統合で データの価値を引き出す

効果的なデータ統合は、組織がデータドリブンのインサイトを獲得、活用する上で必須の要素です。AWS は、ゼロ ETL 統合といった機能や AWS Glue などのサービスにより、これまで煩わしく反復的な作業であったデータ処理プロセスを簡素化しています。このような変化は、チームのオペレーションを改善し、差別化要因としてデータから価値を引き出せる、より生産性の高い環境づくりにつながります。

データがどこにあっても迅速かつ簡単に接続して活用できるように AWS は **ゼロ ETL** の未来に向けた投資を続けています。さまざまなデータソースの統合を AWS がどのように支援しているのか、詳細をご確認ください。

今すぐ見る >

