



データクラウドの導入により データウェアハウスを 最新化すべき5つの理由

データから最大限の価値を抽出する



TABLE OF CONTENTS

- 3** 御社のデータが持つ価値とは
- 4** データウェアハウスが正常に機能していても変更が必要な理由
- 5** 理由1: データウェアハウスのメンテナンスが不要になり、データの有効活用に注力できる
- 6** 理由2: データチームとデータユーザーが制約から解放される
- 7** 理由3: データの質を高めるツールをスムーズに統合できる
- 8** 理由4: コラボレーションが促進される
- 9** 理由5: データの同時アクセス性・実行性が向上する
- 10** ケーススタディ: Snowflakeの導入により次世代薬局サービスを構築したPDX
- 11** データウェアハウスが組織全体に及ぼす影響
- 12** Snowflakeについて

御社のデータが持つ価値とは

今、御社が所有する最も価値の高い資産は恐らくデータかもしれません。ただしそれは、データから詳細なビジネスインサイトを引き出すことができこそです。データは原油と同じように、適切に処理されて使用されることで初めて事業を前に進める価値のあるものとなります。

これは、従来のデータウェアハウスでは実現できません。従来型のデータウェアハウスは、現代の企業が生み出すデータの量、複雑さ、多様性に十分に対応できておらず、リアルタイムにデータにアクセスし分析するという各部門のニーズを同時に満たすことも難しいです。

データクラウドでは、オンプレミスのレガシーなデータウェアハウスや「クラウドウォッシュ」されたデータウェアハウスでは不可能だった、例えば以下のようなことを実現できます。

- すべてのユーザープロファイルとコンテキストデータソースを含む、高度なデータアナリティクスの実行
- ガバナンスや品質を含めたデータ管理とデータストレージの統合
- 管理者の介入を必要としないシステムのコンピューとストレージの自動拡張

- パーティション分割の煩わしさとダウンタイムからの解放
- バックアップスクリプトやデータ保持プロセスのほとんどが不要になる
- チューニングや低レベルのメンテナンス作業から解放される
- 基盤インフラの管理が不要になる

データクラウドでは、企業が保有するあらゆるデータから最大限のビジネスインサイトと価値を引き出すことができ、さらにシステム構成のコストや継続的なメンテナンス、および性能向上のための継続的な調整を最小限に抑えることができます。「組織内の全メンバーが、適時に使いやすい方法で、すべてのデータから価値を引き出すことができる」、これが目指すべき姿です。

データウェアハウスが正常に機能していても 変更が必要な理由

データウェアハウスを最新化するメリットは理解できても、新しいソリューションへの移行に不安を感じるのはごく自然なことです。すでにかかなりの時間と予算をかけてオンプレミスシステムを導入しており、そしてそれは意図したとおりに機能しているかもしれません。不具合もないのに何故変更しなければならないのでしょうか？その理由は以下の通りです。

元々データウェアハウスは複数のソースから収集されるデータ (OLTP) の集中リポジトリとして開発され、様々なビジネスアナリティクス (OLAP) と組み合わせて使用されていました。マーケティング、営業、製品開発、財務などの機能別チーム、および技術、事業部門、役員などのロールを持つユーザーは、ウェアハウス内の常駐データから実用的なインサイトを引き出すことができました。しかし現在では、インフラ的観点から言えば数年前のデータウェアハウスでさえすでに今日の組織のニーズを満たすことはできません。

一方、最新のクラウドデータプラットフォーム技術は従来のデータウェアハウスよりはるかに多くの種類のデータやデータソースの利用に対応しており、これには構造化および半構造化データ、また企業アプリ、モバイルアプリ、ウェブアプリ、APIなどのデータソースも含まれます。すべての情報を統一された方法で結びつけることで、得られるインサイトが飛躍的に向上し、ベーシックなアナリティクスから得られる成果を超えたメリットを享受できます。Snowflakeクラウドデータプラットフォームは、まさにそのような最大限のメリットをデータから引き出すことのできるソリューションです。

次に、データの利用方法を最新化する5つの理由をご紹介します。



理由1: データウェアハウスのメンテナンスが不要になり、データの有効活用に注力できる

バリューチェーン上で真に価値のあるデータドリブンなインサイトや戦略に注力できるのがデータクラウドの強みです。Snowflakeプラットフォームの導入によりウェアハウスインフラの手動での作成や維持管理が不要になり、よりビジネスニーズに適した形でデータを有効利用できます。このようなイノベーションは組織の業務のあり方を大きく変え、システムメンテナンス作業よりも組織としてのより大きな成功に向けたタスクに優先的に取り組むことができるようになります。

Snowflakeプラットフォームはアジリティと経済性というメリットを実現し、必要に応じて手動操作なしにコンピュータとストレージを自動的にスケールアップします。処理容量を自動的にオンライン/オフライン化することで、コンピューティングと人的リソースという特にコストのかかるリソースを最大限に活用できます。こうしたメリットは、規模や製品、サービスを問わずどのような組織でも享受可能です。

タスク	従来のオンプレミス DW	従来のクラウドDW	データクラウド	
🔄 インフラストラクチャ	ハードウェア	自組織	クラウドベンダー	ビルトイン
	ソフトウェア	ディストリビューション	クラウドベンダー	ビルトイン
	ハードウェアクラスタの設定	自組織	自組織	ビルトイン
	ソフトウェアのプロビジョニング	ツール	ツール	ビルトイン
🔒 データとサービスの保護	データの保護と維持	プラットフォームまたはアドオン	プラットフォームまたはアドオン	サービスとしてのデータウェアハウス
	ノード不具合時の保護/リカバリ	自組織	自組織	
	災害からの復旧	自組織	自組織	
	サービスの監視とアラート	自組織	自組織	
🛡️ セキュリティ	物理的セキュリティ	自組織	クラウドベンダー	
	導入セキュリティ	自組織	自組織	
	セキュリティモニタリング	自組織	自組織	
⚙️ データベースの管理と微調整	コンピュータのスケールリング	自組織	自組織	
	インデックス管理	自組織	自組織	
	データパーティション分割	自組織	自組織	
	メタデータおよび統計の管理	自組織	自組織	
	クエリ最適化	自組織	自組織	

図1: Snowflakeなどのクラウドデータプラットフォームは、従来型のクラウドまたはオンプレミスソリューションのような手動のインフラ関連作業を必要とせず、優れたコスト効率でデータから最大限の価値を引き出します

理由2: データチームとデータユーザーが 制約から開放される

データクエリの実行という定常的なユーザーニーズへの対応が常にボトルネックとなってきました。これは、クエリを要求するのが事業部門の責任者や役員であっても、またデータの場所がオンサイトであってもクラウドであっても、データチームが間に入って必要なデータビューやデータスキューを提供する必要があったためです。このため十分な情報に基づいた

迅速な意思決定ができず、新製品の市場投入までの時間が長期化し、顧客のための価値実現までに時間がかかっていました。

しかしSnowflakeのようなクラウドデータプラットフォームを利用すれば、どのチームも幅広いデータと関連するアナリティクスにスピーディーにアクセスできます。Snowflakeのユニークなアーキテクチャは、パフォーマンスを損なうことなく複数のワークロードの同時実行を可能にします。

さらに、Snowflakeでは、スクリプトを書く必要もデータチームにサポートを求める必要もないため、スキルレベルに関係なくどのユーザーも自分のやりたいデータ操作をセルフサービスで実行できます。データ操作はそれぞれのユーザーが使い慣れたSQLなどの言語を使用して迅速かつスムーズに行うことができます。データチームは不要にはなりません、データアクセスポリシーの策定や運用などのより重要なタスクに注力できるようになるでしょう。



理由3： データの質を高めるツールをスムーズに 統合できる

データの質の高さは、計画や戦略を策定するのに不可欠な要件です。これはデータウェアハウスのインフラストラクチャがレガシーベースであっても、最先端のものであっても変わりません。しかしクラウドデータプラットフォームでは意思決定に使用されるデータが多くなるため、データの質の高さがより重要となります。

データの信頼性を保証するにはまずデータ品質の維持が求められます。「garbage in, garbage out (ゴミからはゴミしか生まれぬ)」という格言は、まさにクラウドデータプラットフォームに当てはまる言葉です。クラウドデータプラットフォームの範囲とユーザー層の幅広さを考えると、重要なのは次の2点です。

- 組織にとって本当に重要なデータは何かを判断し、データの優先順位付けや検証を各事業部門やそれぞれの分野のエキスパートに任せられるようにするには、まずデータガバナンスが確立されている必要があります。
- データウェアハウスに入るすべてのデータを、正確かつ利用可能で真のビジネスインサイトを抽出できるものにするには、データをクリーンアップする機能が求められます。

そして、これらの実現に利用できるツールがあります。Snowflakeは複数の専門ベンダー（TalendやInformaticaなど）と密接に連携し、データインポート操作の円滑性を保証する一方で厳格な監視と優れた精度を維持しています。私たちが目指すのは、ユーザーができるだけ幅広いデータにアクセスできるようにし、かつデータがユーザー組織の厳しい基準に適合するよう保証することにより、Snowflake機能に基づいて各種ダウンストリーム処理を最大限に活用していただくことです。



理由4： コラボレーションが促進される

前述のように、データガバナンスとデータの質の高さは、現代のデータプラットフォームに不可欠な要件です。その実現にはITチームの多大な貢献が必要ですが、求められるのは単に技術的対応ではありません。最も正確なアナリティクス結果を得るためには、そのデータについて熟知した現場の実務チームがデータのガバナンスと品質保証に関与する必要があります。しかし、ITチームと実務チームは目に見えない厚い壁で隔てられていることが少なくありません。データを管理しシステムに入力する人（ITチーム）は通常、データを使用する人（実務チーム）のニーズを詳しく把握しておらず、その逆もまた然りです。チーム間のコラボレーションが有益であることは皆が認識しつつも、実際の業務プロセスにそのようなコラボレーションが組み込まれていることはまれです。

このようなジレンマを解消するには、ITチームや実務チームを含む組織全体のデータ管理機能を統合してコラボレーションを促進し、相互依存的なタスクの遂行に必要なツールを各チームに提供する必要があります。

こうした観点から、クラウドデータプラットフォームは以下を実現します。

- データの修正・変更および信頼性確保を含むデータ品質管理機能を確立する。
- データに基づくペルソナや製品特性の定義を含むマスターデータ管理機能を提供する。
- それぞれのデータが正しいコンテキストで利用されるよう、「データに関するデータ」、つまりメタデータを的確に管理および利用する。
- クラウド対応技術に基づいて、チーム間のシームレスなコラボレーションを促進するセルフサービスアプリを提供する。



理由5： データの同時アクセス性・実行性が向上する

旧式のデータウェアハウスにおける制約の1つが、ピーク時のシステム可用性とパフォーマンスを維持しつつ時間的制約の少ないタスクにも対応するためにジョブのキューイングを行う必要がある点です。キューイング型システムでは通常、「モンスタージョブ」は夜間にスケジューリングされ、データローディングなどの操作は通常業務時間のビジネスインテリジェンス (BI) 環境に影響を与えないようオフタイムに実行されます。しかしほぼ誰もが、リアルタイムもしくはそれに近いタイミングで評価や意思決定が行えるよう自分のデータをすぐに処理してほしいはずです。つまりこうした「キューと遅延」に基づくデータ処理方式は誰にとっても決してベストなものではありません。

クラウドデータプラットフォームにはこのようなスケジューリング上の制約はありません。Snowflakeでは、ワークロードが同じタイミングで重複しても、各ワークロードをそれぞれ別のデータウェアハウスに割り当てることができます。これは、アナリティクス情報が即時に利用可能であることが必須な金融業や小売業においては特に重要です。具体的には、Snowflakeは以下を実現します。

- 複数のデータウェアハウス(それぞれ独立したコンピュータクラスター)を無限に作成できる
- すべてのチームが正確かつ整合性のあるデータを共有できる
- システムのパフォーマンスや可用性を損なわない
- どのユーザーもデータのライブビューを即時にクエリできる
- それぞれのユーザーが必要なときに必要な情報を利用して業務を遂行できる

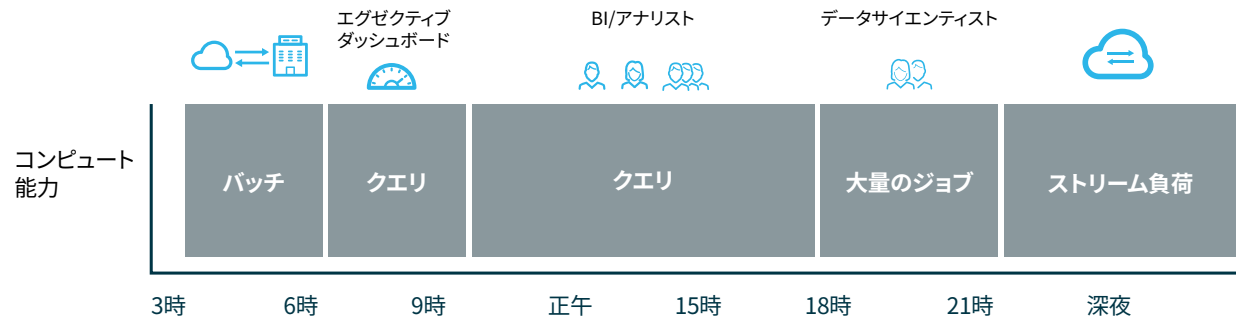


図2: 通常のデータウェアハウス環境では、法人システムのバッチ処理はオフタイムに行われ、ダッシュボードやBIリソースは通常の営業時間中に、大規模なジョブは夜間に実行されます。データのローディングも、生産稼働時のクエリに影響を与えないようオフタイムに実行されます。しかしこのような「キューと遅延」方式のデータ処理システムでは、重要なステークホルダーがタイムリーにデータにアクセスできなくなります。

データクラウド: 別個のワークロードを同時に操作

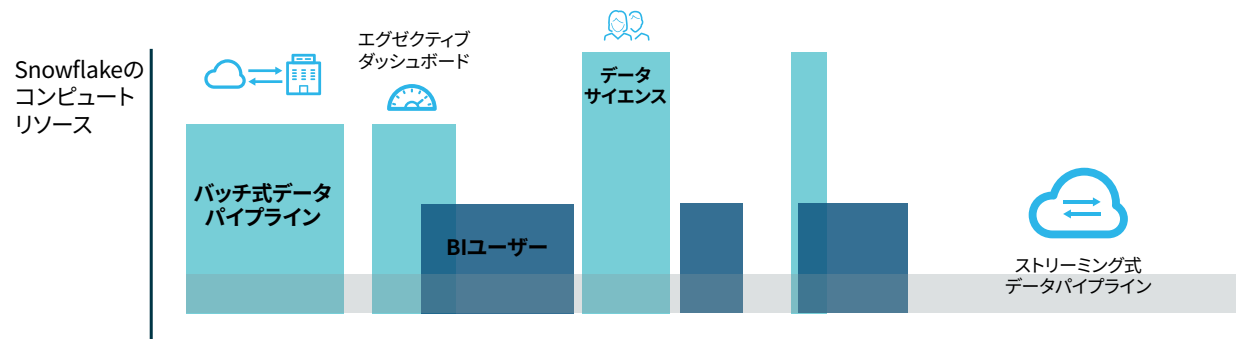


図3: データクラウドは、複数の異なるワークロードをそれぞれ別のデータウェアハウスに割り当てることができます。同時にワークロードを処理してもリソースの枯渇やパフォーマンスの低下が生じないため、意思決定に重要な情報をタイムリーに提供できます。

ケーススタディ：SNOWFLAKEの導入により次世代薬局サービスを構築したPDX

PDXは米国トップクラスの薬局事業者向けヘルスケア技術プロバイダーで、その製品とサービスは、薬局での処方箋処理、医療記録の保存と保護、患者の予後の改善、および厳格に規制された業界でのコンプライアンスおよび競争力の保証に活用されています。PDXは、ユーザー企業へのベストな患者プロフィールの提供を維持しつつ自社システムの収益性向上を希求していました。

PDXは、Snowflakeの導入により、本eBookで紹介するベストプラクティスに基づいた最新データプラットフォームを構築しました。PDXの新しいシステムは、同社の従来のシステムでは困難であった以下のようなメリットを実現しました。

- インフラストラクチャの取得コスト、管理投資、手動によるチューニングが不要になり、ハードウェアや人員の増強を伴わずデータリソースをダイナミックにスケーリングできる。
- 顧客のニーズに合った商品を提供することでデータの価値を最大限に引き出すことができる。
- 複数の薬局店舗や薬局チェーンをまたいだ包括的な販売データの閲覧とアナリティクスをスムーズに取得できる。
- 従来のシステムで発生していたコンピュータ関連の遅延がなくなり、数十億件のレコードを数分で処理することができる。
- ユーザーは生産システムに影響を与えることなくレポートを作成でき、また毎日のBIワークロードをデータ読み込みタスクと同時に処理できる。

「私たちは、どのような状況に直面にしてもSnowflakeソリューションのスケーリングにより薬局事業者様の性能要件を満たすことができると確信しています。」

John Foss氏

PDX、ビジネスインテリジェンス&メーカーレポート担当ディレクター

⁹ Harvard Business Review Analytic Services. An Inflection Point for the Data-Driven Enterprise (データドリブン型企業のための変曲点) <https://www.snowflake.com/resource/an-inflection-point-for-the-data-driven-enterprise/>

データウェアハウスが組織全体に及ぼす影響

Snowflakeのようなクラウドデータプラットフォームの必須要素である最新のデータウェアハウスで、低レベルなデータウェアハウスインフラタスクの排除による時間の節約、システム全体の合理化、コストの削減を実現します。また面倒な手作業やオーバーヘッドがなくなり、システムの性能や可用性が向上します。

クラウドデータプラットフォームは、データ品質の維持、機能チーム間でのコラボレーションの促進、セルフサービスでのデータアクセス性の強化などの業務上のメリットに加えて、全社的にこれから何を目指し達成できるかのビジョンをチームに提供します。

さらに最も重要なことは、組織のバリューチェーン上で「より完全かつクリエイティブなデータの解釈と利用により実効性のある成果が得られる」という、恐らく現状での想定を超えたレベルにデータの価値を高めていくことができます。データ戦略の強化により、鮮度・精度が高くしかも即時に利用可能な情報に基づいた新しい製品やサービスの企画立案が可能になります。そして、技術チームからビジネスマネージャーや役員に至るまで全てのメンバーがそれ

ぞれの役割を推進していくことができます。

今こそデータウェアハウスをアップグレードし、クラウドデータプラットフォームソリューションの全社的メリットを最大限に享受すべき時です。





SNOWFLAKEについて

Snowflakeは、Snowflakeのデータクラウドを用い、あらゆる組織が自らのデータを活用できるようにします。お客様には、データクラウドを利用してサイロ化されたデータを統合し、データを発見してセキュアに共有し、多様な分析ワークロードを実行していただけます。データやユーザーがどこに存在するかに関係なく、Snowflakeは複数のクラウドと地域にまたがり単一のデータ体験を提供します。多くの業界から何千ものお客様（2021年10月31日時点で、2021年Fortune 500社のうちの223社を含む）が、Snowflakeデータクラウドを自社のビジネスの向上のために活用しています。詳しくは、[snowflake.com](https://www.snowflake.com)をご覧ください。



© 2022 Snowflake Inc. All rights reserved.