

学びを簡単に

Snowflake特集号

# データクラウド

for  
**dummies**<sup>®</sup>  
A Wiley Brand

サイロ化したデータを統合する

管理されたデータに世界中から安全にアクセスして共有する

多様な分析ワークロードの実行

提供：

 snowflake<sup>®</sup>

David Baum

## Snowflake社について

Snowflakeは、何千もの組織がほぼ無制限の拡張性と同時実行性でデータを活用しているグローバルネットワークであるデータクラウドを提供しています。データクラウド内では、サイロ化したデータを統合したり、管理されたデータを簡単に見つけて共有したり、多様な分析ワークロードを実行したりできます。Snowflakeは、データやユーザーの場所にかかわらず、複数のパブリッククラウドにわたって1つのシームレスなエクスペリエンスを提供します。Snowflakeのプラットフォームは、データクラウドを支え、データクラウドへのアクセスを提供するエンジンであり、データウェアハウス、データレイク、データエンジニアリング、データサイエンス、データアプリケーション開発、データシェアリングのためのソリューションを作り出します。Snowflakeのお客様、パートナー、データプロバイダーは、すでにデータクラウドでビジネスの新境地を開拓しています。 <https://www.snowflake.com/?lang=ja>



# データクラウド

Snowflake特集号

David Baum著

for  
**dummies**<sup>®</sup>  
A Wiley Brand

# データクラウドFor Dummies、Snowflake特集号

出版元：

**John Wiley & Sons, Inc.**  
111 River St.  
Hoboken, NJ 07030-5774  
www.wiley.com

Copyright © 2022 by John Wiley & Sons, Inc., Hoboken, New Jersey

1976年米国著作権法第107条または第108条によって許可される場合を除き、本書のいかなる部分も、出版社から事前に書面による許可を得ることなく、いかなる形態または電子的、機械的、複写、録画、スキャン等のいかなる手段によっても、複製、検索システムへの保存、または転送することを禁じます。出版社への許可の申請は、Permissions Department (John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, +1 (201) 748-6011, fax +1 (201)748-6008) まで、またはオンライン (<http://www.wiley.com/go/permissions>) で送付してください。

商標：Wiley、For Dummies、Dummies Manのロゴ、Dummies.comおよび関連するトレードドレスは米国またはその他の国におけるJohn Wiley & Sons, Inc.および関連会社の商標または登録商標であり、書面による許可なしに使用することはできません。SnowflakeおよびSnowflakeのロゴはSnowflake Inc.の商標または登録商標です。その他の商標はすべて、それぞれの所有者に帰属します。John Wiley & Sons, Inc.は、本書に記載されているいかなる製品またはベンダーとも一切関係がありません。

責任の制限/保証の免責：出版社および著者は、本書の内容の正確性または完全性に関して何らかの表明または保証をするものではなく、具体的には特定用途への適合性を含めたあらゆる保証を免責されるものとします。販売または販売促進資料により、保証が発生または延長されることはありません。本書に記載する助言および戦略は、あらゆる状況に適するものではありません。本書は、出版社が法務、会計、またはその他の専門サービスの履行に従事していないという理解のもとに販売されます。専門的な補助が必要な場合は、資格を持つ専門職によるサービスを要請してください。出版社および著者のいずれも、本書に起因する損害に対して責任を負わないものとします。組織またはWEBサイトが本書で引用または詳細情報の提供源として参照されているという事実は、著者または出版社がかかる組織またはWEBサイトが提供する情報または推奨事項を公認していることを示すものではありません。さらに読者は、本書に記載されたインターネットWEBサイトは、本書の作成時から現時点までの間に変更または消去されている可能性があることを承知するものとします。

弊社のその他の製品およびサービスに関する全般的な情報、または読者の皆様のビジネスまたは組織向けのカスタムFor Dummiesブックを作成する方法については、弊社の事業開発部門（米国、877-409-4177、[info@dummies.biz](mailto:info@dummies.biz)）までお問い合わせいただくか、[www.wiley.com/go/custompub](http://www.wiley.com/go/custompub)をご覧ください。製製品またはサービスへのFor Dummiesブランドのライセンス付与については、[BrandedRights&Licenses@Wiley.com](mailto:BrandedRights&Licenses@Wiley.com)までお問い合わせください。

ISBN 978-1-119-86754-8（ペーパーバック）；ISBN 978-1-119-86755-5（電子書籍）

製作：アメリカ合衆国

10 9 8 7 6 5 4 3 2 1

## 謝辞

本書ならびにその制作に携わってくださった方々は弊社の誇りです。本書の出版にあたりご協力いただきました皆様に心より御礼申し上げます。

開発編集者：Brian Walls

プロジェクト責任者：  
Martin V. Minner

シニアマネージングエディター：  
Rev Mengle

アキジションエディター：  
Ashley Coffey

事業開発担当：William Hull

プロダクションエディター：  
Mohammed Zafar Ali

Snowflake寄稿者：Vincent Morello、  
Elise Bergeron、Kent Graziano、  
Tim Fletcher、Clarke Patterson、  
Ganesh Subramanian、  
Christina Jimenez、  
Leslie Steere

# 目次

はじめに .....	1
Snowflakeのデータクラウドを導入する .....	1
本書で使用するアイコン .....	2
本書を読み終えた後で .....	2
第 1 章 <b>データによって生み出される課題           と機会を把握する</b> .....	3
膨大で多様なデータに対処する .....	4
サイロに屈すればサイロが増える .....	4
分断されたデータの問題を解決する .....	5
データガバナンスに注意を払う .....	6
クラウドを積極的に導入する .....	7
サイロを打破する .....	8
データクラウドの影響を理解する .....	9
クラウドネットワークを介してデータを共有する .....	10
今後の展望 .....	12
第 2 章 <b>データクラウドの価値と能力を理解する</b> .....	13
データクラウドで実現できることを理解する .....	13
データにアクセスする .....	14
データを管理する .....	14
データをアクションにつなげる .....	14
データクラウドのユニークな特性を把握する .....	15
1つのデータクラウドに標準化する .....	16
あらゆるデータをサポートする .....	16
あらゆるワークロードを強化する .....	16
データ共有の可能性を広げる .....	17
Snowflakeプラットフォームのご紹介 .....	18
クラウドに適した拡張性とパフォーマンス .....	18
抜群の経済的価値 .....	18
固有の使いやすさ .....	19
マルチクラウドとクロスクラウドの柔軟性 .....	19
組み込みのセキュリティ .....	20
独自のコラボレーションオプション .....	20

第 3 章	<b>データクラウドでコラボレーションを行う</b> .....	21
	データの再帰率を理解する.....	22
	最新のデータ共有方法の紹介.....	22
	組織の4つの壁を超える.....	24
	Snowflakeデータマーケットプレイスを活用する.....	25
	Snowflakeデータマーケットプレイスで差別化する.....	27
第 4 章	<b>各業界におけるデータクラウドの展開</b> .....	29
	金融サービスの価値を高める.....	29
	医療分野で好結果を生み出す.....	32
	小売業サプライチェーンを強化する.....	34
	優れたメディアおよびエンターテインメントサービスを提供する.....	38
	公共サービスの質を向上させる.....	41
第 5 章	<b>SNOWFLAKEのプラットフォームを掘り下げる</b> .....	43
	適切なアーキテクチャを使用して開始する.....	43
	過去の教訓を活かす.....	44
	パフォーマンスを高めコストを削減する.....	45
	適切なアーキテクチャが重要な理由を理解する.....	46
	マルチリージョン、マルチクラウドを1つのサービスで.....	47
	使いやすいプラットフォームを享受する.....	48
	使用量の予測と監視.....	48
	あらゆるタイプのデータへ容易にアクセス.....	49
	強力なセキュリティとガバナンスを適用する.....	50
第 6 章	<b>あらゆるワークロードを実行する</b> .....	53
	データウェアハウスとデータレイクの導入.....	53
	コアワークロードの強化.....	54
	その他の重要なワークロードをデータクラウドで実行する.....	55
	データパイプラインのエンジニアリング.....	55
	データサイエンスを簡素化する.....	56
	データアプリケーションの構築.....	56
	データを制限なしに共有する.....	58
第 7 章	<b>データクラウドの利用を開始するための 6つのステップ</b> .....	59

# はじめに

**イ**ノベーターはデータを使ってビジネスを構築し変革を起こします。そのために、データを簡単かつ安全に統一、統合、分析し、自社のエコシステム内で、および同じ目的を持つ他の組織と共有することを可能にするテクノロジーを求めています。

残念ながら、このようなデータの大部分はサイロで作成され、そこにとどまります。データは、オンプレミス、クラウド、ソフトウェアアプリケーション、顧客接点のどこで生成されようと、部門、データセンター、パブリッククラウドに分断されます。サプライチェーンのオペレーション、POS トランザクション、データ保護対策アプリ、その他多数のビジネスプロセスによって独自のデータストアが生み出されます。このようなデータにアクセスし、管理し、ビジネスに活用するのは容易なことではありません。

## Snowflakeのデータクラウドを導入する

データクラウドは、何千もの組織がほぼ無制限の拡張性、同時実行、性能でデータを活用しているグローバルネットワークです。データクラウド内では、サイロ化したデータを統合したり、管理されたデータを簡単に見つけて安全に共有したり、多様な分析ワークロードを実行したりできます。データクラウドは、データやユーザーの場所にかかわらず、またデータやワークロードが複数のパブリッククラウドにまたがって存在するときでも、統一されたシームレスなエクスペリエンスを提供します。データクラウドは、データを見つけて管理し、事業部門、サプライヤー、ビジネスパートナー、顧客との間で共有できます。さらに、Snowflakeデータマーケットプレイスでは、125社を超えるパートナーが提供するデータやデータサービスにライブでアクセスすることもできます。さまざまな業界でほぼ無限の機会があります。

- ▶ **小売業者**は、データクラウドを利用してライブデータを簡単に一元化し、CPG（消費財）企業やサプライチェーン、その他のパートナーと共有して、価格を最適化し、在庫回転率を高め、利益の増大を目指します。
- ▶ **金融サービス企業**は、プロセスをデジタル化・自動化し、不正行為やリスクを軽減し、セカンドパーティデータやサードパーティデータに安全にアクセスし、それらを自社データと組み合わせることにより、価値の高い顧客サービスを提供します。
- ▶ **医療機関**は、データクラウドを利用することにより、ライブの医療データを組織内及びパートナーと安全に共有することで、患者の治療結果を改善し、コストを削減し、市場投入までの時間を短縮することができます。

- ▶▶ **メディア企業**は、加入者データを一元化し、ブランド、広告主、広告プラットフォーム、広告制作会社と共有することで、加入者の生涯価値、広告収入、ROI（投資収益率）を高めることができます。
- ▶▶ **製造業者**は、データクラウドを利用することにより、自社のビジネスエコシステム全体でサプライチェーンの活動を同期させ、工場の生産性を高め、データドリブなパートナーネットワークを構築しながら製造品質を向上させることができます。
- ▶▶ **公共分野の組織**は、ITを最新化し、機関、政府、パートナー間でデータを安全に共有してコラボレーションを行うことで、新しいインサイトを獲得し、市民サービスを改善することができます。

## 本書で使用するアイコン

本書では、以下のアイコンを使用してヒントや覚えておくべき重要ポイント、実例などを強調しています。



ヒント



ポイント



ケーススタディ



技術情報

組織でデータクラウドをどのように活用できるかをアドバイスします。

データクラウドの理解を深めるにあたり、覚えておくべき概念を示しています。

データクラウドを利用して自社のデータを統合、共有、活用している組織の事例をご紹介します。

専門用語を詳しく解説しています。

## 本書を読み終えた後で

本書に興味を持っていただけましたら、<https://www.snowflake.com/?lang=ja>にアクセスしてください。データクラウドの無償トライアル版の注文、プランや価格に関する詳細情報の入手、ウェビナーを視聴することができます。また、詳細なドキュメントにアクセスし、Snowflakeの担当者に連絡することもできます。

- » データの多様性を理解する
- » サイロ化したデータの問題を解決する
- » データクラウドに情報を一元化する
- » データエコシステムの可能性を理解する

## 第1章

# データによって生み出される課題と機会を把握する

ビジネス上のやり取りでは、それがソーシャルメディア、モバイル通信、IoT（モノのインターネット）デバイス、eコマース取引、またはさまざまなタイプのデジタルサービスのどこで行われようと、ほぼ間違いなくデータが生成されます。このようなやり取りに関わる人やデバイス、やり取りが行われる場所が増え続けていることを踏まえると、その規模は途方もない大きさになります。そしてそれが毎日繰り返されるのです。その結果、ビジネスの世界では、データを保存・管理する必要性がこれまで以上に高まっています。

最近のForbesのレポートによれば、世界中で作成、捕捉、コピー、消費されたデータ量は、2010年の1.2兆ギガバイトから2020年には59兆ギガバイトへと大きく増加しています。IDCは、2020年から2023年までの3年間に作成されるデータ量が、過去30年間に作成されたデータ量を上回ると予測しています。MarketsandMarkets社の調査では、この新たに流入する膨大な量のデータに対応することを一因として、世界のクラウドコンピューティング市場の規模が2025年までに現在の2倍を超える8,320億ドルに達すると見込まれています。

その一方で、Accenture社が750名の事業経営・IT担当上級専門職を対象に行った調査では、クラウドの活用によって期待した成果を完全に達成できたと答えた回答者は37%に過ぎませんでした。また、クラウドに移行することで予定期間内に意図した価値を実現できると確信していると答え

た回答者はわずか29%でした。本章で説明するようにこのような、データによって生み出される課題と新たな可能性に対応するための新しいパラダイムが求められています。それに応えるのがデータクラウドです。

## 膨大で多様なデータに対処する

ビジネス上の課題となるのはデータの膨大な量だけではありません。データの多様性もまた課題となります。IDCのレポート「2020 Global DataSphere (2020年度Global DataSphere)」によると、新たに作成されるデータの約80%が半構造化データ（ウェブログ、IoT、モバイルデバイスのデータなど）または非構造化データ（オーディオ、ビデオ、PDF、その他のタイプのリッチメディアコンテンツ）です。エンタープライズコンピューティングのバックボーンである従来型データベースは、オンプレミスであれクラウドであれ、こうした大量の、または多様な情報を一元化、統合、分析、共有できるようには設計されていません。

このような短所により、多くのビジネスリーダーは、グローバルなデータの供給、需要、消費という新しいデータエコノミーに参加するにはどうすればよいかと頭を悩ませています。有用なデータの総量や種類が増え続ける中、価値の実現までの時間を短縮するという要求は高まるばかりです。残念ながら、データ管理や分析用ソリューションのほとんどは、利用可能なデータのごく一部しか活用できず、将来ではなく過去に目を向けています。こういったシステムは依然として重要性の高いものではありませんが、今日企業に必要なのは、事業の推進、最適な予測に基づく顧客への最良のサービスの提供、マーケットの新たなチャンスの発見に役立つ予測的・規範的なインサイトをもたらすテクノロジー、データ、グローバルネットワークに簡単にアクセスできることです。

確かなことが1つあります。データ中心の運用はもはやオプションではありません。データを一元化し、他の何千もの組織から標準化されたシームレスな方法でデータやデータサービスにアクセスできる機会を設けることが、これまではあり得なかった規模で成功を収めるための最新の手法として急速に普及しつつあります。

## サイロに屈すればサイロが増える

ITチームはこれまで、技術的な制約によりデータをソフトウェアやネットワークの境界の背後に隔離せざるを得ませんでした。そのため、新しいデータ管理の取り組みを開始するたびに別のデータサイロが作成されました。当初、このようなリポジトリはデータウェアハウスとそれを補完する

### 4 データクラウドFor Dummies、Snowflake特集号

データマート、データレイクという形を取っていましたが、最近では、機械学習やデータサイエンス向けに設計された新しいタイプのデータベースが数多く登場しています。その結果、世界のデータはデータタイプ、ワークロード、リージョン、そしてクラウドによって分断されています。多くの組織はこのような差異を調整するのに苦労しており、多くのサイロに存在するデータは結合して分析するのが非常に困難です。

ユースケースによっては複数のソースからのデータが必要になる場合があります。たとえば、Hadoopデータレイクにある半構造化データと1つ以上のデータウェアハウスにあるリレーショナルデータを結合しなければならない場合です。IT担当者は、こうした多様なソースからのデータを分析するために、独自の高価なハードウェアによる複雑なデータ取り込み手順を必要とする専用のデータウェアハウスを構築しなければならない場合があります。データエンジニアは、このようなデータセットを統合する手段として、各ソースからデータのサブセットを抽出し、それをまた別のサイロにマージするカスタムコード化手順を作成します。

このような脆弱なインターフェイスは、新しいデータソースやデータターゲットに対応するために継続的な更新を必要とします。たとえば、データサイエンスアプリでは通常、ビジネスインテリジェンスアプリとは異なる形式（またはモデル）でのデータ管理が必要になります。このような差異を調整し、すべてのデータを同期するには、膨大なプログラミング作業が必要になります。



ポイント

高度な分析アプリケーションや機械学習モデルの多くは、個別のデータセットを活用しています。異なるソースにまたがって分析するのは非常に難しいためです。

## 分断されたデータの問題を解決する

仮に、組織の全員が、一貫性のある管理されたデータが格納されている1つの共通のリポジトリにアクセスできるとしましょう。あらゆる種類のデータのあるシステムから別のシステムにインポートまたはエクスポートすることなく結合することがどれほど簡単になるか想像してみてください。経営幹部、マネージャー、業務担当者が組織全体で信頼できる唯一の情報源を活用できるとしたらどうなるでしょうか。ソフトウェア開発者、データサイエンティスト、データエンジニア、ビジネスアナリストが、分析するデータの準備に費やす時間を減らし、データ分析自体にもっと多くの時間をかけることができれば、組織の生産性はどれほど向上するのでしょうか。

このような「たられば」は、40年近くにわたって悩みの種のようにソフトウェア業界にまわりついてきました。その主な原因は、企業情報シ

システムにおけるデータの保存方法とデータへのアクセス方法にあります。オンプレミスであれクラウドであれ、各本番アプリケーションには、データを保持しておく固有の場所や形式があります。以下に例を挙げてみましょう。

- ▶▶ マーケティングデータはマーケティングオートメーションシステムに保存される。
- ▶▶ 販売データは顧客関係管理（CRM）システムに保存される。
- ▶▶ 財務データはエンタープライズリソースプランニング（ERP）システムに保存される。
- ▶▶ 在庫データは倉庫管理システムに保存される。

本番データを分析用に抽出すると、業務報告用のデータウェアハウス、部門別分析用のデータマート、データマイニングおよび探索用のデータレイクなど、また別のサイロが作成されます。このようなデータ管理システムでは、データを読み込んで分析の準備をするために専用のETL（抽出、変換、ロード）ツールが必要になります。



技術情報

多くの組織は、データベースおよびコンピューティングプラットフォーム間でのデータ交換を調整するデータパイプラインを設定するために高い費用をかけてソフトウェアエンジニアを雇っています。また、データタイプ間の差異を合理的に処理し、レポート、分析、データサイエンス用の新しいターゲットデータベースを作成するための専用の統合ツールも購入しています。データにアクセスする手順やデータを結合、マージする手順は複雑で、管理コストが高くつき、拡張も容易ではありません。

## データガバナンスに注意を払う

データウェアハウスやデータレイクでのデータ保存にベストプラクティスを採用している場合でも、セキュリティとガバナンスに関しては長年の課題があり、年々厳しくなるデータプライバシーの各種規制によりさらに複雑化しています。たとえば、EU（欧州連合）域内で事業展開する企業であれば、GDPR（一般データ保護規則）の要件を満たすために、データ系列およびトレーサビリティに関する厳格な要件の遵守が求められます。米国カリフォルニア州でも同様のCCPA（カリフォルニア州消費者プライバシー法）という規制が施行されています。ヘルスケア分野にはHIPAA（医療保険の相互運用性と説明責任に関する法律）、eコマース分野にはPCI DSS（クレジットカード業界情報セキュリティ基準）、金融分野にはSOX（サーベンス・オクスリー法）など、業界特有の規制があり、セキュリティとガバナンスは複雑化する一方です。



ポイント

データを複製するたびに、新しいサイロに政府や業界の規制を適用しなければなりません。すべての複製データとデータインスタンスでデータの証跡をたどる必要があるため、データサイロの数が増えるほど、コンプライアンスの確保は複雑になります。このようなサイロを統合する最善の方法は、データを単一のリポジトリに統合し、外部データへのシームレスかつ高パフォーマンスなアクセスを実現することです。



ヒント

適切なガバナンスが欠如していると、データサイロは、データ系列の追跡、データのカタログ化、セキュリティルールの適用がさらに困難になる原因となり、企業がコンプライアンスに取り組む上での妨げになります。一元化されたリポジトリにデータを統合すると、このような作業が簡素化されます。

## クラウドを積極的に導入する

多くの組織は、大量のデータを保存および共有するため、Amazon Web Services (AWS)、Microsoft Azure、Google Cloud Platform (GCP) などのパブリッククラウドのリポジトリにデータを置いています。これらのユビキタスサービスはデータを保存、共有、収益化するための道を開きました。IDCが2020年9月に発表した「Quarterly Cloud IT Infrastructure Tracker (四半期クラウドITインフラ追跡)」によると、2019年から2020年にかけて、企業のパブリッククラウドおよびプライベートクラウドに対する支出は34.4%増加したのに対し、クラウド以外のIT関連支出は8%減少しました。

IDCやその他調査機関が指摘しているように、近年はクラウドを採用する傾向が着実に高まっていますが、新型コロナウイルス感染症の世界的大流行で働く環境が一新されたことに伴い、2020年にはクラウドコンピューティングとクラウドストレージの利用がさらに勢いを増しました。自宅で働く従業員へのサポートにより、データストレージとデータ管理関連の活動が急激に押し上げられたためです。リモートワークが当たり前になり、遠隔地とのコミュニケーションやデータの共有が簡単に行えるアプリケーションが注目を浴びるようになりました。

当初、このようなパブリッククラウドサービスは、増え続けるデータサイロの解決の一翼を担いましたが、十分な容量があるからといってデータアクセスの問題が解決するとは限りません。このような潤沢な容量は、多種多様なパブリッククラウド、クラウド間に特有の非互換性、さまざまなサブスクリプションモデルにより、かえって混乱を引き起こし、差異の拡大を招く可能性があります。そのため、クラウドは新たなデータサイロの発生源となっています。その主な理由は、主要なクラウドサービス間でデータやワークロードを統合することが容易ではないことです。今やクラウドデータベースは簡単に立ち上げられますが、従来のオ

ンプレミスの世界と同様に、そうしたデータベースはすぐにそれ自体のサイロになってしまう可能性があります。

Deloitte社による2020年12月の「Cloud Migration Forecast（クラウド移行予測）」レポートによると、ITマネージャーの97%が、レジリエンスを最大限に高めること、規制・コンプライアンス要件を満たすこと、およびさまざまなプロバイダから最善の組み合わせのサービスを選んで活用することを目的に、複数のクラウドにワークロードを分散することを計画しています。

このようなパブリッククラウドのリソースが簡単に利用できるようになると、データの多様性の問題がさらに大きくなります。なぜなら、個人がクレジットカードを使用して（通常はIT部門の管理が及ばない場所に）データベースを立ち上げることができるからです。このような部門のデータベースと情報システムは適切に展開、バックアップ、保護、統合されているとは限らず、データの適切な配布、保護、利用について明確に規定したITポリシーに準拠していない可能性もあります。

パブリッククラウドコンピューティングの急成長と歩調を合わせて膨大な数のSaaS（サービスとしてのソフトウェア）アプリケーションが生まれ出され、今も増え続けています。このようなSaaSアプリケーションには、それぞれ独自のデータ形式とリポジトリがあります。現在、多くの企業は、この巨大な「アプリケーションクラウド」への投資の調整に苦心しています。大企業では、このようなアプリケーションクラウドに数百から数千もの独自のアプリやデータストアが含まれる場合があります。



ポイント

パブリッククラウドサービスは、データの保存やアクセスを容易にしますが、ベンダーのクラウド間にある固有の非互換性により、新しいデータサイロの発生源となっています。これらのクラウド間ではデータ資産を容易に共有または移動できないため、データセットの複製が増加し、データ管理の悩みの種が増えます。

## サイロを打破する

誰もがデータを結合し、そのサービス提供対象者（顧客、見込み客、市民、患者、その他の関連グループ）を360度の角度から完全に把握したいと考えています。たとえば、マーケティング担当者は通常、顧客と見込み顧客1人1人の全体像を把握するため、利用可能なあらゆるタッチポイントを通じて顧客とのやり取りを記録し、電子メール、電話、チャット、Webサイトへの訪問、ソーシャルメディアへの投稿、店頭取引、カスタマーサービスとのやり取りからデータを引き出しています。場合によっては、マーケティングチームはセカンドパーティデータやサードパーティデータを収集しています。

このような重要な360度ビューを達成するには、これらすべてのデータを1か所に集めておく必要があります。たとえば、最も有望なリードを生み出す広告キャンペーンを特定するため、マーケティング担当者はCRMデータ、コールセンターデータ、マーケティングキャンペーンデータを結合し、顧客1人1人のユニークな「ジャーニー」を把握しようとする場合があります。それ以外の部門の組織も同様の課題に直面しており、政府部門では「市民360」、医療部門では「患者360」という言葉が生まれています。

いずれの場合であれ、完全なインサイトを得る上での障害は、異なる種類のデータを収集し結合させることの難しさにあります。どの場合においても、異なる種類のデータを収集して結合することの難しさが完全なインサイトを獲得する上での障害になります。たとえば、医療機関は、医療記録システムにある構造化データと構造化されていない手書きの診断書、X線画像やMRI画像などの半構造化データを結合しなければならない場合があります。法律の世界では、訴訟に関する大量の文書を、幅広い自由形式の検索機能を備えた形で保持する必要があります。



ポイント

共通のリポジトリにデータを置くと、顧客をセグメント化し、顧客の接触相手、顧客が使用しているチャネル、顧客が関心を持っているオファー、それぞれの製品、サービス、キャンペーンに最も効果的なコンテンツなどのトレンドを把握することが簡単にできるようになります。

医療、法律、マーケティング、財務、その他いずれの分野においても、ビジネスプロフェッショナル、データアナリスト、データエンジニア、データサイエンティスト、アプリケーション開発者は、レポート、分析、データサイエンスといった取り組みで一貫した成果をもたらすために、信頼できる唯一の情報源に自信を持ってアクセスする必要があります。データクラウドはこのようなエクスペリエンスを可能にします。データが複数のクラウドにある場合でも、あるシステムから別のシステムにデータをインポートまたはエクスポートすることなく、すべてのデータを同時に活用できます。このアーキテクチャは、特定のワークロードや単一のデータタイプ用に最適化されていたこれまでのデータアプリケーションの作成方法とはまったく対照的です。

## データクラウドの影響を理解する

データクラウドは、複数のパブリッククラウドにまたがるグローバルなデータネットワークです。データクラウドでは、組織が利用しているクラウドサービスに関係なく、組織全体がどこからでも安全かつシームレスに管理された方法でデータにアクセスし、分析や共有を行うことができます。

データクラウド内では、サイロ化されたデータを統合したり、管理されたデータを簡単に共有したり、各種のデータ駆動型ワークロードを実行

できます。すべてのデータが1か所に集約されているため、データの保守に伴うサイロ化やそれに関連する管理業務を排除できます。



ポイント

データは、データクラウドの一部として、データクラウドエコシステム内の他のデータと関連付けられることで価値が高まります。たとえば、データクラウドでは、結合された一連のデータサービスを活用しながら、管理されたデータをパートナーと共有することが簡単にできます。

データクラウドは、以下のようなソフトウェア業界の最も差し迫った課題を解決することで、デジタルエコノミーにおける企業のニーズに応えます。

- ▶▶ 広範なデータ駆動型ワークロードをサポートする
- ▶▶ 多様なデータソースと広範なデータコンシューマーを結び付ける
- ▶▶ コストパフォーマンスに優れたパブリッククラウドサービスを活用する

## クラウドネットワークを介してデータを共有する

デジタルエコノミーでは、どのような分野の企業にも、データの共有が必要となります。たとえば、小売業者は、販売データをベンダーと共有して在庫とサプライチェーンを管理し、通信事業者は、加入者のエンゲージメント情報をアプリケーション開発企業と共有してカスタマーエクスペリエンスの向上に役立てています。

データクラウドは、データコンシューマー、データプロバイダ、データサービスプロバイダがシームレスで透明性の高い方法でデータとデータサービスを交換可能にすることで、このような相互作用を促進します。たとえば、一部の企業は、データを結集してその価値を最大限に引き出せるようにするサービスを提供しています。このようなデータサービスプロバイダを利用すれば、データが持つ可能性を活性化・管理し、それを十分に理解することで、ビジネス価値を高めることができます。また、データ統合プロバイダはデータをデータクラウドに簡単に取り込めるよう支援します。データサイエンスおよびアナリティクスパートナーはデータから価値を引き出せるよう支援し、システムインテグレーションはデータを安全に共有して収益化できるよう支援します。

データクラウド内では、データプロバイダは小売業者が消費者向け包装商品（GPG）企業とデータを共有するような1対1の共有関係や、気象局が加入者と天気予報データを共有するような1対多の共有関係を構築できま

す。第3章で説明するように、データクラウドからはSnowflakeデータマーケットプレイス経由で何百もの商用データリストにアクセスすることもできます。このマーケットプレイスは、データクラウドのコンポーネントの1つで、豊富なデータセットやデータサービスを提供しています。



ヒント

データクラウドは、新しいグローバルデータシェアリングアーキテクチャを基盤とし、あらゆるユーザーに、このフェデレートされたランドスケープへの参加を可能にします。許可されたユーザーは、データを別の場所に



ケーススタディ

## 新型コロナウイルス感染症のデータを共有することで世界的大流行を軽減

新型コロナウイルス感染症の医療データと患者データへのアクセスを効率化するため、ヘルスケアとテクノロジーの両分野のトップ企業数社が合同でかつてない規模の医療データリポジトリを立ち上げました。「COVID-19 Research Database（新型コロナウイルス感染症研究データベース）」と名づけられたこの業界横断的コンソーシアムは、データクラウドに格納され、学術、科学、医学分野の何千という研究者と共有されています。このリポジトリには、患者の請求データ、電子医療記録（EHR）、米国の死亡データなど、匿名化された数十億件の記録が格納されています。

医療費請求、薬剤費請求、臨床検査データ、人口統計データ、その他多数のデータポイントに関する最新のデータを確保することは、研究者や公衆衛生担当者が致死率の高い新型コロナウイルス感染症に対する理解を深め、対策を練る上で非常に有用です。数十の医学部と数百の公衆衛生機関の許可されたユーザーがこの膨大なクラウドデータセットを活用し、各州のロックダウン政策が新型コロナウイルス感染症対策以外の医療サービスに与える影響、事業再開を検討する際の経済と健康のバランスの取り方、新型コロナウイルス感染症が予防医療に与える影響、各地域の検査コストなど、さまざまな研究を行っています。研究者は、人工呼吸器を必要とする患者の属性要因や既往症を確認したり、隔離政策やソーシャルディスタンス政策の効果を調べたりすることもできます。

セキュアなデータシェアリングテクノロジーにより、これらの研究者はデータをコピーしたり移動したりすることなく共同作業を行うことができます。許可されたユーザーは組織内と複数の組織間の両方でデータを直接共有できます。また、成長を続けるデータマーケットプレイスを利用し、データプロバイダやサービスプロバイダから提供されるすぐに照会可能なライブデータにアクセスすることもできます。

コピーしたり移動したりすることなく、データクラウドで利用できるすべてのデータにアクセスできます。データへのアクセスは、普遍的なデータガバナンスモデルに従って所定の場所で行われます。

## 今後の展望

あらゆる種類のデータの生成と消費の増加に後押しされ、世界のデータスフィアは成長を続けています。現在、ほとんどの組織は、ビジネスの内外で利用可能なデータのごく一部にしかアクセスできていません。データの保存、処理、管理を目的に設計されたテクノロジーを基盤とする新たなパラダイムが求められています。データクラウドは、適切なテクノロジーをもたらし、何千もの組織がデータを結集させ、ビジネスサービスに活かすことを可能にします。第2章では、データクラウドの構造とデータクラウドが組織にもたらすメリットについて説明します。



ポイント

データクラウドは、何千もの組織が自分たちのデータと増え続けるデータセットやデータサービスにほぼ無限の規模と同時性をもって、簡単かつ安全にアクセスできるグローバルネットワークです。データクラウド内では、サイロ化されたデータを統合したり、管理されたデータを簡単に見つけて安全に共有したり、さまざまな分析ワークロードを実行したりできます。データクラウドは、複数のパブリッククラウドにまたがってシームレスなエクスペリエンスを実現します。

## 本章の内容

- » データクラウドで実現できることを理解する
- » データクラウドのユニークな特性を把握する
- » 主要なデータクラウドワークロードを確認する
- » 最新型データシェアリングを探る
- » 基盤となる技術プラットフォームのご紹介

# 第2章 データクラウドの価値と 能力を理解する

データクラウドによって、ビジネスリーダー、データエンジニア、データサイエンティスト、ビジネスアナリスト、オペレーションスタッフはデータをさまざまな方法で活用し、自社のビジネスや顧客に最高のサービスを提供できます。本章では、データクラウドの主要な機能をご紹介しますとともに、そのグローバルデータネットワークの持つユニークな特性に焦点を当てます。

## データクラウドで実現できることを理解する

Snowflakeがデータクラウドを開発した目的は、あらゆる組織がデータサイロ（オンプレミスとパブリッククラウドの両方）を解消し、多様な分析ワークロードを実行し、子会社、エコシステム、リージョンの間で管理されたデータを簡単かつ安全に共有できるようにすることです。主なメリットは、「アクセス」、「ガバナンス」、そして「アクション」の3つのカテゴリーに分類されます。

## データにアクセスする

データクラウドは、構造化データ、半構造化データ、特定のタイプの非構造化データを統合（この機能は今後のリリースで提供予定）するデータプラットフォーム、データウェアハウス、データレイクを個別に用意する必要をなくします。また、データクラウドは、組織全体およびパートナー、サプライヤー、顧客、その他の関係者からなる外部エコシステムとのデータの安全な共有プロセスを簡素化します。さらに、データクラウドの一部であるSnowflakeデータマーケットプレイス経由でデータセット、データサービス、およびデータアプリケーションにアクセスすることで、エコシステム内でデータを共有するためにコピー・移動することなく分析を強化できます。

## データを管理する

データクラウドを利用すると、データサイロを排除し、ビジネスにとって重要なあらゆるデータを管理できます。柔軟なデータガバナンスやセキュリティポリシーを策定して実施し、さまざまなデータワークロードに適用できます。Snowflakeは重要なデータプライバシー規制に対応しているため、セキュリティまたはコンプライアンス違反のリスクを最小限に抑え、イノベーションを妨げることなくデータガバナンスの取り組みを効率化できます。データアクセスおよびデータセキュリティポリシーは、パートナーのエコシステム全体およびサードパーティプロバイダーからアクセスされるデータに拡張できます。さらに、Snowflakeデータマーケットプレイスを経由したサードパーティのデータガバナンスサービスの利用も可能です。

## データをアクションにつなげる

データクラウドは、データエンジニアリング、データレイク、データウェアハウス、データサイエンス、データアプリケーション、データシェアリングなど、あらゆるデータワークロードでほぼ無限のパワー、拡張性、同時実行性を実現できるよう設計されています。データクラウドでは、データを分析したり、新しいデータ製品を開発したり、機械学習モデルを実行して情報に基づいてビジネス上の意思決定を下したりできます。また、新しいデータドリブなインサイトを活用してイノベーションを推進する、データを収益化して新しい収益源を生み出すといったことも可能です（図2-1参照）。

## データクラウド：グローバルネットワーク

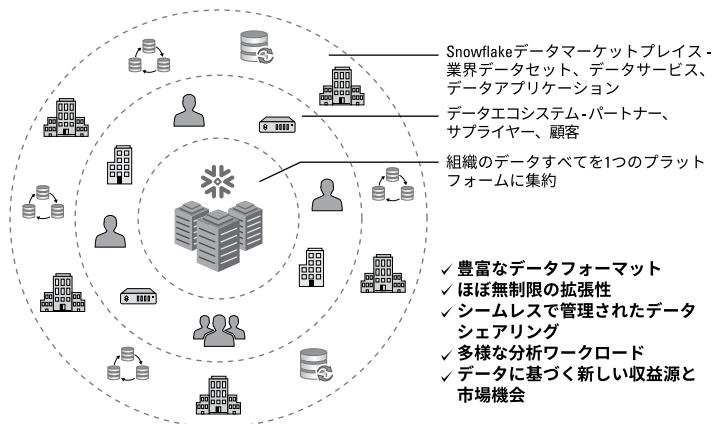


図2-1：データクラウドは、画期的なテクノロジー、データの移動やコピーが必要ないシームレスなデータアクセス、そして多くのビジネス上のメリットをもたらします。

## データクラウドのユニークな特性を把握する

データクラウドは、プラットフォームとデータの2要素で構成されています。Snowflakeのプラットフォームは、データクラウドを動かすエンジンにあたります。データは、顧客と他のデータプロバイダーから提供されます。顧客とデータプロバイダーは、プラットフォームを介してデータの統合、アクセス、そして最終的には共有を行います。

あらゆる規模の、そしてさまざまな業界の組織がデータクラウドに自社のデータを集約し、多様な分析ワークロードを効率的に実行しています。データクラウドでは、同じデータセットに対して複数のワークロードを同時に実行できます。たとえば、データエンジニアリングワークロードをデータウェアハウスおよびデータサイエンスワークロードと同時に実行する必要がある場合などです。



技術情報

ストレージとコンピュート機能を分離するSnowflake独自のアーキテクチャにより、実質的に無制限の数のワークロードがサポートされます。

さらに、データクラウドでは、ファイルをコピーしたり、データサイロを作成したりすることなく、制御された方法でデータとサービスを共有できます。たとえば、データクラウドを使用して、サプライチェーン全体のデータを統合したり、ビジネスパートナーとのやり取りを簡素化し

たり、管理されたデータを収益化して同業他社と共有することで、全く新たな収益源を生み出したりできます。

## 1つのデータクラウドに標準化する

データクラウドは、ある意味でWWW（ワールドワイドウェブ）に似ています。ウェブは、人々が情報を共有したり、文書をリンクしたりできる「情報スーパーハイウェイ」を提供するために作り出されました。データクラウドも同様です。誰でも自由に接続して自分のデータにアクセスしたり、データを管理・有効化したりできます。さらに、より大きなデータエコシステムに接続することもできます。これには、Snowflake データマーケットプレイスのサードパーティデータやデータサービスへのアクセスが含まれます。

データクラウドは、4大陸の3つの主要なパブリッククラウドにまたがり、リージョン間でのデータおよびワークロードの容易なコラボレーションを可能にする複数のSnowflakeリージョンで構成されています。すべてのデータクラウドワークロードは一貫した方法で実行され、ユーザーが利用するリージョンやクラウドプロバイダーに関係なく、すべてのユーザーに1つの一貫性のあるエクスペリエンスが提供されます。

## あらゆるデータをサポートする

2014年、Snowflakeは、構造化データと半構造化データの両方をサポートする最高クラスの単一プラットフォームを導入し、データ管理と分析の世界に変革をもたらしました。また、音声、動画、テキスト、画像データ、PDF、その他多くの種類のファイルを含む特定の種類の非構造化データのサポートを将来のリリースで追加する予定です。データクラウドは、事業活動で生成されたデータと機械が生成したデータの両方を、それぞれのネイティブフォーマットで保存することができます。このような幅広いデータストレージとデータアクセス機能により、データクラウドは「管理されたデータへの無制限のアクセスを提供することで、あらゆる組織が現在の課題と機会に取り組むとともに、将来の可能性を明確化可能な世界という高い理想を実現できます。

## あらゆるワークロードを強化する

データクラウドは、複数のデータベース管理システム、データエンジニアリングツール、クラウドベンダーの管理に伴うコスト、複雑さ、オーバーヘッドなしであらゆるデータをビジネスサービスに活用できるようにします。ですが、データクラウドの真の価値は、データの格納、管理、保護を容易にするだけにとどまりません。データクラウドはデータへのアクセスも効率化します。

データソリューションでは、データを「入れる」のは簡単なことです。しかし、データからインサイトを取得することは複雑になる場合があります。たとえば、分析ワークロードとデータサイエンスワークロードを同時に実行する場合などです。データの取得から利用までを管理できる強力なSnowflakeプラットフォームを基盤とするデータクラウドは、広範なワークロードをサポートすることに秀でています。図2-2に主要なデータクラウドワークロードを示します。詳細については、第6章をご覧ください。

### 1つのデータクラウドで多くのワークロードに対応



図2-2: データクラウドは、世界で最も一般的なデータ管理ワークロードと分析ワークロードをリソースの競合なしにシームレスにサポートします。

## データ共有の可能性を広げる

第1章で触れたように、データクラウドは、何千もの組織が迅速にデータを結集させて共有できるグローバルネットワークを実現します。データプロバイダーは、データを移動したりコピーしたりする必要なく、管理されたデータを簡単にデータコンシューマーと共有できます。コンシューマーは、共有データを発見し、アクセスし、独自のデータと統合して、新たなインサイトを生み出すことができます。



ポイント

共有できるのはデータに限りません。データサービスプロバイダーは、データサービスを介して独自のソリューションやビジネスロジックを共有できます。たとえば、データサイエンスサービスのプロバイダーは、データセット全体へのアクセスを提供せずに、顧客が自社のデータに対して実行できる事前学習済みの機械学習モデルを提供することができるでしょう。

## 高度なデータ共有

データクラウドは、次のような方法で、広範でグローバルなエコシステムでデータを共有および管理するプロセスを合理化します。

- データをコピーすることなく、データおよびデータサービスを共有可能。
- 承認済ユーザーは、常に最新のデータセットにライブで読み取り専用アクセスが可能。
- データコンシューマーは、共有データへのアクセス権限を取得し、自らのデータと組み合わせて新たなインサイトを生み出すことができる。
- Snowflakeデータマーケットプレイスでは、サードパーティプロバイダーの豊富なデータセットやデータサービスが提供される。

## Snowflakeプラットフォームのご紹介

Snowflakeは創業以来、データ管理、データ運用、データ分析の多くの側面の再検討を重ね、クラウドコンピューティングの潜在能力を活用してきました。Snowflakeのプラットフォームはデータクラウドの機能を強化します。このセクションでは、プラットフォームの主な特性に焦点を当てます。Snowflakeプラットフォームを支える技術については、第5章で解説いたします。

### クラウドに適した拡張性とパフォーマンス

Snowflakeプラットフォームは、データ量、コンピュートパフォーマンス、ワークロードの同時実行など、クラウド規模のコンピューティングに対応するよう構築されています。この独自のクラウド向けアーキテクチャによって、データクラウドを利用する組織は、スケールや同時実行の制限なく新しいワークロードを立ち上げ、それらのワークロードに必要なだけのコンピュート機能やストレージ容量を使用してワークロードを動的にプロビジョニングできます。隠れた容量制限も、他のデータワークロードとリソースの競合が発生する心配もありません。

### 抜群の経済的価値

Snowflakeの費用対効果の高いユティリティモデルでは、データクラウドをご利用の組織にお支払いいただくのは、使用容量分の料金（秒単位計算）のみです。容量はユーザーが設定したしきい値に基づいて自

動でプロビジョニングされるため、お客様はデータの利用に集中でき、必要なリソースはプラットフォームで管理されます。Snowflakeがパブリッククラウドのベンダーとの契約で大規模に大量の容量を確保しているため、お客様にはSnowflakeを介してこのようなクラウドサービスを非常に安価にご利用いただけます。

## 固有の使いやすさ

Snowflakeのプラットフォームは、前世代のクラウドデータプラットフォームやDBaaS（サービスとしてのデータベース）に比べ使いやすさが格段に向上しています。プラットフォームの実行に専門知識が不要なため、データベース管理者は手動でのチューニングや設定作業から解放され、データモデルの作成と改良、新しいインサイトの抽出、広範なデータからビジネス価値を引き出すことに集中できます。

## マルチクラウドとクロスクラウドの柔軟性

データクラウドは、世界で最も普及しているクラウドサービスであるAmazon Web Services (AWS)、Microsoft Azure、Google Cloud Platform (GCP) すべてに対応します。これらのいずれかに基づいて標準化する必要はもうありません。ビジネスニーズに応じて複数のリージョンやクラウドを利用でき、あるリージョンやクラウドから別のリージョンやクラウドにデータをシームレスに共有できます。たとえば、マルチクラウド戦略を採用する多国籍大企業であれば、データクラウドを利用して複数のクラウドプロバイダーリージョンにまたがってシームレスに運用し、世界中の顧客ベースにサービスを提供し、災害復旧対策を強化することができます。同様に、データプロバイダーであれば、Snowflakeのグローバルインフラを活用して、さまざまなリージョンで、複数のクラウドにまたがって他の組織とデータを共有することができます。



ポイント

Snowflakeは基盤となるパブリッククラウドのプラットフォームを完全に抽象化しているため、ユーザーや管理者にクラウド特有の専門知識は必要ありません。ユーザーや管理者は、このようなサードパーティのクラウドサービスを直接操作する必要はまったくなく、パブリッククラウドの使用料金を別途請求されることもありません。



ポイント

Snowflakeでは、プラットフォームをクラウドに依存しない設計を採用しているため、AWSでもMicrosoft AzureでもGCPでもまったく同様にプラットフォームが動作します。統一された1つのコードベースが3つのパブリッククラウドサービスすべてに及んでいるため、お客様は1つの一貫性のあるプラットフォームインターフェイスを使用したまま、それらのクラウド間でシームレスにデータとワークロードを移動できます。クラウドに依存しないというのは、世界中の顧客ベースにサービスを提供したり、グローバルな災害復旧戦略を策定したりする際に決定的なメリットとなります。

## 組み込みのセキュリティ

データから分析価値を引き出すためには、強力なデータガバナンスとコンプライアンスの枠組が不可欠です。データクラウドでは、すべてのデータを格納したまま、そのデータへの高速かつ管理された安全なアクセスを提供することができます。すべてのデータとワークロードに業界をリードするサイバーセキュリティ対策が適用されます。マルチレベルのセキュリティ機能には、データが保管されているすべての施設を保護する物理的なセキュリティ、データ暗号化とそれに関連するキー管理サービス、ロールベースのアクセス制御、オブジェクトレベルの権限、頑丈なデータベースセキュリティなどがあります。さらに、データクラウドでは、規制やデータプライバシーポリシーを遵守し、機密性の高いデータの適正な取り扱いを保証しています。

多くの分野で新しいプライバシー要件が生まれ続けていますが、Snowflakeプラットフォームは、お客様が国や業界特有の規制を遵守し、規制違反によるコストを回避できるよう、進化を続けています。

## 独自のコラボレーションオプション

Snowflake独自のアーキテクチャは、組織全体およびエコシステム全体にわたってデータを活用したコラボレーションに取り組んでいる組織にほぼ無制限の拡張性、同時実行性、パフォーマンスを提供します。アクセスするのがデータセットであろうとデータサービスであろうと、コラボレーション相手がエコシステム内の他の組織であろうと、従来のデータプロバイダーやSaaSベンダーであろうと、データクラウドは可能性のレベルを引き上げます。サプライチェーン全体のデータを統合して、市場投入までの時間を短縮することができます。また、まったく新しい収益源を確立したり、新しく生産的な方法でビジネスパートナーとコラボレーションすることもできます。Snowflakeのお客様は、自社のデータの枠を超えることで、豊富なインサイトを獲得し、優れた製品を作り出し、最高のサービスを提供することができます。

- » 時代遅れのデータシェアリング方法を刷新する
- » 組織の内外とデータを共有する
- » Snowflakeデータマーケットプレイスを活用する
- » データエコノミーの進展に足並みをそろえる

## 第3章

# データクラウドでコラボレーションを行う

**頑** 固なデータエコノミーを形成する上での主な障害の1つは、データの共有が難しく、コラボレーションの妨げとなっていることです。組織内のデータ所有者は、さまざまなデータプラットフォーム、データベースモデル、クラウドにデータをサイロ化して格納していることがよくあり、データの交換や拡散が複雑なものになります。生データは一般に異なるフォーマット（構造化、半構造化、非構造化）で存在しており、統合やアクセスを容易に行うことができません。さらに、データガバナンスの手法やデータの使用方法に関する規制は地域や会社ごとに異なっており、一般データ保護規則（GDPR）、カリフォルニア消費者プライバシー法（CCPA）、その他多くの業界のデータプライバシー規制への準拠が複雑になります。

データクラウドでは、次のようなSnowflake独自の機能を活用してこのような問題を解消することができます。

- » Snowflakeセキュアデータシェアリングは、顧客がデータを移動させることなく、組織内外でデータをリアルタイムで安全に共有できるようにします。
- » Snowflakeマーケットプレイスは、データコンシューマーがサードパーティのデータセットやデータサービスに自社のデータと同様にアクセスしてクエリを実行し、これまで想像もつかなかったようなビジネスインサイトを見つけ出せるようにします。

# データの再帰率を理解する

今日のビジネスに影響を及ぼす重要な要因の1つとして、ユニークデータ（新規に作成・取得されるデータ）と複製されたデータ（コピーされて利用される既存データ）の比率があります。IDCの2020 Global Data-Sphereレポートによると、この比率は1対9で、1対10に近づきつつあります。言い換えれば、コピーされて利用されるデータは、新規に作成されるデータよりも10倍多いということです。この現象はデータの再帰率とも呼ばれ、同じデータが繰り返し処理される割合を示します。

このような統計値の持つ意味は、ほとんどのデータ共有方法に伴うコスト、手間、リスクを踏まえるといっそう明確になります。最も一般的な方法は、データソース間をプログラムでリンクすることです。たとえば、FTP（ファイル転送プロトコル）サイトの開設やAPI（アプリケーションプログラミングインターフェイス）の作成などです。これらはいずれもコードの作成と保守を必要とします。他には単にデータのコピーをメールでやり取りする、ファイル共有ユーティリティでコンテンツを共有するといった方法がありますが、このような単純な方法はどれもデータセキュリティや規制ポリシーに準拠していません。



ポイント

コピーされて共有されたデータはすぐに陳腐化するため、常に更新が必要になります。これがデータの再帰率が高いこと的主要原因となっています。通常は、データの移動が1回で済むことはありません。データセットは最新に保つ必要があり、そのためには定期的に一括アップロードして以前のデータセットを置き換えるか、個別のトランザクションを実行して変更を記録する必要があります。

## 最新のデータ共有方法の紹介

データクラウドは、前述のような従来型のデータ共有方法を刷新するSnowflakeのセキュアデータシェアリングテクノロジーを採用しています。セキュアデータシェアリングでは、古くてメンテナンスの手間がかかる手動のソフトウェア手順は必要ありません。代わりに、アクセス権限を設定するだけで、複数の組織がデータクラウドで安全にライブデータを共有し、利用することができます。データプロバイダーは、対象となるデータコンシューマーと共有したいデータへのアクセスを簡単に許可できます。データをコピーまたは移動したり、手動で更新したりする必要はありません。

すべてのデータベースオブジェクトは、データクラウド内で一元的に保持、更新されます。クエリやトランザクションが複数のパブリッククラウドにまたがって実行される場合でも、エンドツーエンドのセキュリティ、ガバナンス、メタデータ管理サービスが体系的に適用されます。アプリケーションを連携したり、ファイル共有手順を設定したり、データを最新に保つために新しいデータを頻繁にアップロードしたりする必要はありません。又、コストを低く抑えることもできます。データはコピーされるのではなく共有されるため、ストレージを追加する必要がないからです。

Snowflakeのお客様は、Snowflakeアカウント内の特定のデータセットへの管理されたアクセスを許可するだけでデータを取得・共有することができます。受取側のアカウントは別のクラウドにあっても同じクラウドの異なるリージョンにあっても構いません。プロバイダーは共有するデータを指定して、そのデータに権限を付与します。受け取り側ではそのデータを物理的に保存する必要はなく、すぐに照会できます。

データクラウド内の広範なガバナンスとセキュリティによって、データプロバイダーはSnowflakeアカウント内でのデータ管理を維持しながら、パートナーと安全にデータを共有することができます。データプロバイダーは、1対1、1対多、多対多のデータシェアリングリレーションを作成できます。データプロバイダーはセキュリティとコントロールを維持しながら、データクラウドの他のメンバーとデータの一部を共有することができます。このアプローチの利点は明確です。つまり、データは常に最新の状態で、レイテンシはほぼゼロで、コンシューマーは共有データセットへの更新をすぐに利用できます（図3-1を参照）。

### データクラウドによる最新のデータシェアリング

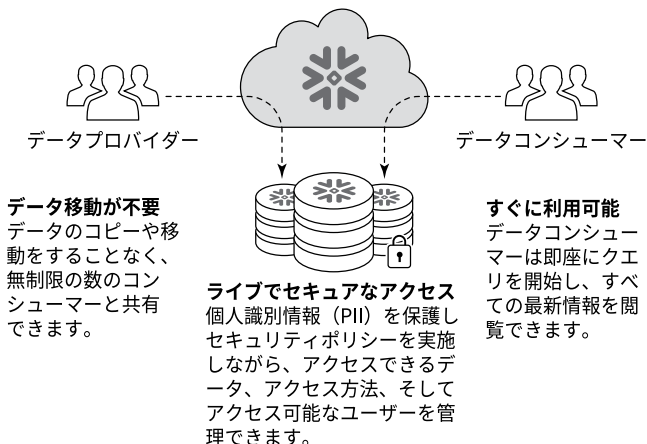


図3-2: データクラウドは、データプロバイダーとデータコンシューマー間のデータシェアリングを、それが複数のリージョンやクラウドにまたがる場合であっても効率化します。

# データシェアリングの進化

データシェアリングにデータクラウドを使うべき理由とは？

- **インターフェイスの簡素化**：データのコピーや移動のために、電子メール、複数の表計算シート、共有ネットワークドライブ、APIなどの従来の方法を使用する必要がなくなります。
- **レイテンシの低減**：不完全な古いインサイトにつながることが多い陳腐化したデータを排除できます。
- **範囲の拡大**：企業全体で、さらには社外の何千もの組織との間で、データを発見し、アクセスし、共有することができます。



ポイント

Snowflakeセキュアデータシェアリングの技術により、複数の組織が簡単にライブで管理されたデータを共有し、利用できます。従来のデータシェアリング方法とは異なり、追加のメンテナンスやファイル転送は必要ありません。自社内、ビジネスエコシステム全体、さらには顧客やパートナーなど、さまざまな組織とデータを共有し、共有データを受け取ることができます。

## 組織の4つの壁を超える

今日のビジネスでは、コラボレーションが1つの組織の枠を超えて行われることがますます増えています。データクラウド内では、顧客、サプライヤー、パートナーなど、ビジネスエコシステム内の他のメンバーと自由にコラボレーションすることができます。データクラウドが持つ共有の可能性によってネットワーク効果が生まれます。つまり、データクラウドに参加する組織が増えるほど、他のSnowflakeの顧客、パートナー、データプロバイダーと交換できるデータが増加し、その結果、すべての利用者にとってのデータクラウドの価値が高まります。

たとえば、出張管理サービスを提供しているソフトウェア会社は、データクラウドを利用してクレジットカードでの購入、経費データ、予算情報を共有することで、顧客が会社の出張経費を分析するのを支援できます。医療情報調査会社は、データクラウドを利用して地方政府、州政府、連邦政府機関と新型コロナウイルス関連のデータを共有することができます。金融サービス会社は、データクラウドを利用して市場トレンドや為替レートに関するデータを公開することで、証券

会社や投資会社が市場の動きを把握するのを支援できます。また、サプライヤーやパートナー、遠隔地の事業部門との連携を望む企業は、データクラウドを利用してコラボレーションを行うことができます（図3-2を参照）。

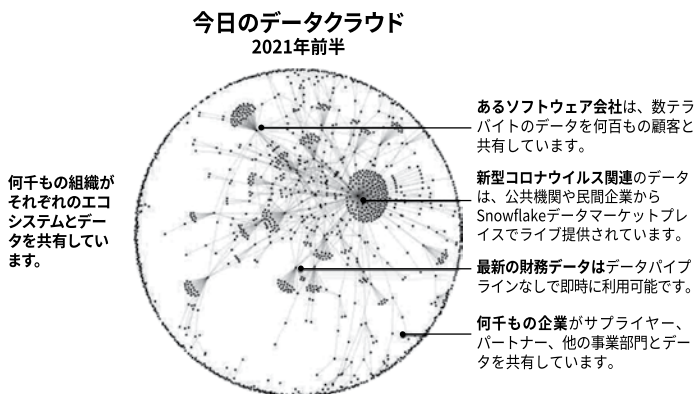


図3-2: データクラウドは、あらゆる種類の企業間、およびあらゆる種類のワークグループ内でのデータ交換を効率化します。

## Snowflakeデータマーケットプレイスを活用する

このような目的をさらに推進し、新たなデータの可能性を拓くため、データクラウドにはSnowflakeデータマーケットプレイスが含まれています。Snowflakeデータマーケットプレイスでは、FactSet、Starschema、Weather Source、S&P Global、Wunderman Thompson、Epsilon、Tapad、AccuWeather、Airlines Reporting Corporation（ARC）など、125以上の商用プロバイダーが提供する何百ものデータセットやデータサービスを利用することができます。

さらに、Snowflakeのお客様は管理されたデータセットをリスト化し、それをSnowflakeデータマーケットプレイス経由で他のSnowflakeの顧客と共有することで、データセットを収益化することができます。データコンシューマーは、サードパーティのデータを活用したり自社のデータと統合したりするための生産的な方法を見つけることができます。データプロバイダーやデータサービスプロバイダーは、新しいオーディエンスに素早くリーチして収益の拡大を図ることができます。

他のデータマーケットプレイスとは異なり、Snowflake独自のデータクラウドアーキテクチャでは、顧客はファイルのコピーやデータの移動を行うことなく、ライブですぐにクエリ可能なサードパーティのデータセットやデータサービスを利用することができます（図3-3を参照）。

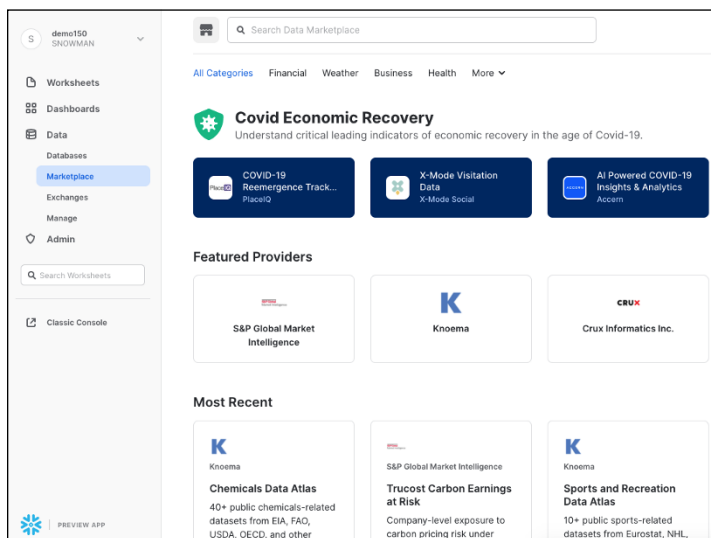


図3-3：Snowflakeデータマーケットプレイスには、12以上のカテゴリにわたる125以上のデータパートナーのデータやサービスが用意されています。

Snowflakeには、このデータを生産的な方法で発見、管理、監視し、共有するための機能があります。承認済のデータコンシューマーは、閲覧を許可されたデータに安全にアクセスできます。たとえば、次のようなデータを利用できます。

- ▶ **ARC**は、234の国と地域の4,600以上の空港からのフライトを網羅する、世界で最も包括的な国際航空旅行情報へのアクセスを提供しています。
- ▶ **S&P Global**は、企業がビジネスパートナー候補の信頼性を判断するのに役立つ格付情報を提供しています。企業はこの情報を活用することで、信用に関するリスクを全体的に把握できます。
- ▶ **Heap**は、Snowflakeデータマーケットプレイスを使用してウェブサイトやアプリケーションからのユーザー、セッション、イベントレベルのデータを共有しています。

Snowflakeのお客様は、不正取引の検出、リスクスコアの評価、セキュリティログに潜むサイバーセキュリティ脅威の特定、マーケティングのための顧客データの拡充などの有用なデータサービスをマーケットプレイスで利用することもできます。たとえば、次のようなデータサービスを利用できます。

- ▶▶ **LiveRamp**は、マーケティング担当者が100以上のオフラインデータソースからデータポイントを取り込み、より完全な顧客プロファイルを作成する支援を提供します。
- ▶▶ **Quantifind**は、財務リスク評価を行い、人や組織のリスクレベルを算定します。
- ▶▶ **Hunters**は、既存のセキュリティソリューションを回避する攻撃を、人工知能によって検知する自律型サイバーセキュリティソリューションです。
- ▶▶ **Seismic**は、マーケティングおよびセールスイネーブルメントデータをその分析レポートデータセットの機能として共有するとともに、ユーザー、コンテンツ、アクションを管理するためのデータモデルを提供しています。



ヒント

データクラウドのお客様は、Snowflakeデータマーケットプレイスを利用して有用なデータセットやデータサービスにアクセスできます。また、自社のデータを収益化することで新たな収益源を確立できる場合があります。Snowflakeデータマーケットプレイスで世界中からアクセスできるデータの例として、人口統計、マーケティング、気候、新型コロナウイルスなどのデータセットがあります。

## Snowflakeデータマーケットプレイスで差別化する

他のデータマーケットプレイスでは、データコンシューマーがデータパイプラインを構築し、データを個別ファイルとして取り込み、オンプレミスかパブリッククラウドのいずれかにある自社のシステムに保存する必要があります。データコンシューマーは、データプロバイダーがデータセットを更新するたびにこの作業を行わなければなりません。プログラムによるリンクまたは手動のファイル転送を使用するカスタムインターフェイスを作成するには、複雑なデータエンジニアリングが必要になる場合があります。

Snowflakeのセキュアデータシェアリングテクノロジーにより、Snowflakeデータマーケットプレイスのユーザーとデータサービスプロバイダー

は、数回クリックするだけでライブデータセットを照会することができます。ファイルを取り込むことなく、各コンシューマーのSnowflakeアカウントから自動的に情報にアクセスできます。追加のストレージ、メンテナンス、ファイル転送は必要ありません。コンシューマーは、データプロバイダーがデータセットを更新すると同時に変更や更新をすぐに閲覧できます。

## SNOWFLAKEデータマーケットプレイスを介してリアルタイムデータを配信する



ケーススタディ

FactSet社は何万人もの投資プロフェッショナルに財務データ、マーケットデータ、分析サービスを提供しています。このような顧客は、アプリケーション、ウェブポータル、統計パッケージなどにデータを統合し、重要な意思決定を行っています。

以前、FactSet社ではデータの転送にAPIコール、FTP、Secure FTPを使用していました。これには時間がかかり、大量の計算リソースが必要でした。非常に大きなデータセットを取り込むのに数週間もかかったことがあります。FactSet社は、顧客にコンテンツを配信するより効率的な方法を模索していました。

現在、FactSet社はデータクラウドにデータセットを一度だけ展開し、Snowflakeデータマーケットプレイスを介してほぼ無制限の数の顧客のデータセットを即座に利用できるようにしています。投資プロフェッショナルは、FactSet社が管理するコンテンツ（20以上の独占データフィードや数十のサードパーティデータフィードを含む）に、ストレージコストを一切かけずに迅速かつほぼリアルタイムでアクセスできます。投資プロフェッショナルは、構造化データや半構造化データをSQLで照会し、自社のデータと結合し、結合されたデータセットを自社のツールで分析できます。煩雑なETL（抽出、変換、読み込み）の手順を使用する必要はありません。

ファイル共有プロセスは、かつては数時間、チームがクラウドリソースをプロビジョニングする必要がある場合は数日間かかることさえありましたが、現在はほぼ即座に行うことができます。FactSetの顧客の一部は、FactSetのクエリと他のコンテンツプロバイダーのデータを結合するなどして、Snowflakeデータマーケットプレイスを利用してバラバラのデータセットをマッピングしています。FactSet社は、米国財務省のTreasury International Capital (TIC) システムの経済データや調査データ、FactSet社の提携企業であるCornerstone Advisors社の銀行ベンチマークデータなど、さらに多くのデータセットをSnowflakeデータマーケットプレイスに追加する予定です。

## 本章の内容

- » 金融サービスの価値を高める
- » 医療およびライフサイエンス産業で好結果を生み出す
- » 最新のインサイトで小売業サプライチェーンを強化する
- » 優れたメディアおよびエンターテインメント体験を提供する
- » 卓越した政府サービスを市民に提供する

# 第4章

# 各業界におけるデータクラウドの展開

**デ**ータクラウドは、さまざまな業界の組織がデータの価値を最大限に引き出し、タイムリーなインサイトの提供を通じて効率性の向上、収益の増大、コストの削減、顧客、患者、市民へのより良い体験の提供を実現することを可能にします。本章では、いくつかの業界を例にデータクラウドの可能性について説明します。

## 金融サービスの価値を高める

今日の金融サービス企業は大きな変化に直面しています。このような変化を引き起こしている要因の1つは、データ、分析、人工知能を駆使して消費者や企業にデジタル金融サービスを提供することに長けた新しいタイプの俊敏な金融テクノロジー（フィンテック）企業の台頭です。PwCは、「Financial Services Technology 2020 and Beyond：Embracing Disruption（金融サービステクノロジー2020年およびその後：ディストラクションを受け入れる）」と題したレポートを発表しました。その中で、今日の金融サービス企業の4分の3以上が、顧客を日常的なコミュニ

ケーションの中心に据え、高度にパーソナライズされた体験を可能にするためにテクノロジーに投資していることを明らかにしています。

データクラウドは、データを取得、保存、分析、共有するプロセスと、価値の高いサービスを生み出す新しいデータアプリケーションを開発するプロセスを簡素化します。金融サービス企業は、データクラウドを利用することで、データを一元管理、データからインサイトを獲得、自社のデータをベンダーや金融機関パートナーからのセカンドパーティデータと統合ができます。また、Snowflakeデータマーケットプレイスなどのソースからサードパーティデータを収集し、顧客や投資家をより幅広い視点でとらえることもできます。たとえば、銀行は、その企業部門、商業部門、小売部門内の複数の接点から顧客データを収集しています。管理されたデータの共有ソースをデータクラウドに置くことで、顧客特有のニーズを満たす新しい製品やサービスを開発するためのベストプラクティスを簡単に特定することができます。もう1つの例は投資会社と資産運用会社です。これらの会社は、最新のデータを使って投資戦略の差別化を図り、アドバイザーチームが迅速にデータ主導型的意思決定を行えるようにしています。

多くの金融サービス企業は、サードパーティデータセットを活用し、自社のビジネスに影響を与える経済的および社会的動向に関して理解を深めています。また、不正行為の検知、マネーロンダリング防止、リスク管理、信用評価などの重要な規制およびガバナンスタスクにデータサービスを活用しています。



ポイント

企業がすべてのデータを一元化され、世界中で利用でき、一貫した方法で管理されるプラットフォームに保存できれば、情報の分析がはるかに容易になります。

保険会社はデータクラウドを使用して、パートナー企業からのサードパーティデータセットや、Snowflakeデータマーケットプレイスで公開されているデータセットにアクセスします。請求漏れを最小限に抑えたい損害保険会社（P&C）は、たとえば、Quantifindデータサービスを利用して、不正の可能性のある自動車修理請求の調査を行うことができます。Snowflakeデータマーケットプレイスで利用可能なこのサービスは、制裁措置やブラックリストなどの外部データソースを、リスクレベルを示す予測リスク類型モデルと組み合わせて使用しています。このデータは、損害査定担当者が、請求を提出する各個人の不正リスクスコアを算出するのに役立ちます。

データクラウドによって、損害査定担当者はサードパーティデータにすぐにアクセスできます。これにより、保険会社のビジネスエコシステム全体でのコラボレーションが促進されます。たとえば、損害保険会社の

役員が、年間支払保険金総額などの財務指標を、同じ地域の同様の指標と比較したいと考えるとします。ここでも、Snowflakeデータマーケットプレイスが必要なインサイトを提供します。SnowflakeのパートナーであるS&P Globalは、National Association of Insurance Commissioners (全米保険監督官協会) を情報源とする包括的な法定財務データを提供しています。

Snowflakeのセキュアデータシェアリングテクノロジーにより、データを転送したりカスタムAPIを設定したりせずに、このような複数のデータソースが即座に利用できるようになります。そのため、ビジネスインテリジェンスやデータサイエンスの取り組みにデータを簡単に活用できます。たとえば、損害保険会社では、Quantifind社のデータとS&P Global社のデータ、さらに社内の顧客管理 (CRM) システムにある各自動車修理工場に関する情報を組み合わせてエグゼクティブダッシュボードを作成することができます。

## 価値の高い顧客および投資家体験

金融サービス企業は、データを集約して自社に関するより深いインサイトを得ること、セカンドパーティデータやサードパーティデータに安全にアクセスして顧客と投資家をより幅広い視点でとらえること、不正行為やリスク発生を減らすこと、およびプロセスをデジタル化・自動化して価値の高いサービスの提供に専念できるようにすることを目的にデータクラウドを利用しています。

- **金融機関**：顧客データを一元化して安全に共有することで、すべての部門でサイロを排除するとともに、企業全体で顧客を360度の視点から把握できるようにします。
- **金融サービスパートナー企業**：顧客の行動データと財務データをフィンテックや他のパートナー企業からの追加のソースと組み合わせることで、より広範な顧客プロファイルを作成します。
- **金融データプロバイダー**：金融機関が顧客との関係を深められるように関連する顧客データと投資データを共有することで、提供する商品とサービスを拡大します。
- **カスタマーエクスペリエンス**：金融サービス企業は、あらゆる接点で顧客、投資家、契約者にタイムリーでパーソナライズされたエクスペリエンスと適切なコミュニケーションを提供できます。

普遍的なデータレポジトリを確保することで、サーベンス・オクスリー法やドッド＝フランク・ウォール街改革・消費者保護法などの金融規制、およびその他多数の国際規制への遵守を可能にする制御を適用しながら、それぞれの事業部門、子会社、データパートナーとのデータシェアリングが簡素化されます。

## 医療分野で好結果を生み出す

医療とライフサイエンスは類似の特性を持つ異なる産業です。医療が患者への医療サービスの提供を中心に据える一方で、ライフサイエンスは薬品や治療法の開発と商業化に重点を置いています。この分野は本質的に非常に重要性が高いため、厳しい規制を受けています。米国では、これらの産業に影響を与える最も有名な医療情報セキュリティ要件は、保護医療情報（PHI）と個人識別情報（PII）の使用法を規定する「医療保険の相互運用と説明責任に関する法律（HIPAA）」です。欧州連合では、医療データは一般データ保護規則（GDPR）の要件に従って保護する必要があります。

データクラウドは、機密性の高い医療データの安全を確保しつつ、他の医療系システム、支払者ネットワーク、研究パートナー、医療情報分析プロバイダーなどヘシームレスで、ほぼリアルタイムに管理されたアクセスを提供します。SnowflakeはHIPAAに準拠し、HITRUSTの認定を受けているため、すべての患者データが迅速かつ安全に、これらの厳格なデータガバナンス基準に従って統合されます。

電子カルテ（EHR）システムやコネクテッドヘルスデバイスから生成される患者の医療データが急激に増加したことで、医療の提供、臨床診断、医療イノベーション、および規制上の意思決定への新たなアプローチの道が開かれています。医療提供者、支払者、ライフサイエンス企業、そして医療研究機関の間で、膨大な量のデータがやり取りされています。データクラウドは、安全なデータ管理機能を提供しながら、日常的な管理報告と高度な分析およびデータサイエンスの取り組みの両方を支援します。

医療機関は、Snowflakeのセキュアデータシェアリング機能を活用することで、保護医療情報を保護し、データセキュリティとプライバシー規制への準拠を維持しながら、ライブデータへのセキュアで管理されたア

クセスを提供して真の相互運用性を実現できます。さらに、Snowflake データマーケットプレイスには、医師の資格認定および医師ベース市場 シェアデータから公衆衛生データまで、医療に特化したデータセットや データサービスもあります。



ポイント

PHIおよびPIIのセキュリティを規定する業界規制への準拠は、多くの医療・ライフサイエンス企業にとって継続的に取り組むべき課題です。しかし、臨床データや管理データの保存、管理、統制方法を再定義できる組織にとっては非常に大きな機会となります。

## 治療プロセス全体を通じて信頼できる 唯一の情報源

医療およびライフサイエンス企業は、質の高い患者転帰をもたらすこと、成長を促進すること、市場投入までの時間を短縮すること、コストを削減すると同時にHIPAAなどのデータガバナンス規制への準拠を可能にするコントロールを導入することを目的に、データクラウドを利用してライブの医療データを組織内やパートナーのネットワークと安全に共有しています。

- **医療提供者**：電子カルテに加え、臨床データ、管理データ、経営データ、医療分野のIoT (IoMT) データなどの患者データの唯一の情報源を作成します。
- **支払者**：保険金請求データや領収書などを医療提供者と共有し、処方集データと診断データを薬局や研究パートナーと共有します。
- **ライフサイエンス企業**：臨床、販売、マーケティング、その他のデータを事業部門内や外部の研究協力者と共有します。
- **研究機関**：医療提供者、ライフサイエンス企業、その他のパートナーから収集された健康転帰データを集約して共有します。



ケーススタディ

## 患者治療のための共有データ

Komodo Health社は、疾病負担を軽減するには、データと分析をよりスマートに、かつ革新的な方法で利用することが不可欠であると考えています。このため、同社は人工知能をはじめとする高度なデータサイエンス技術を、3億2千万人以上の患者一人ひとりの治療過程を追跡する「ヘルスケアマップ」に適用し、分析アルゴリズムと臨床の専門知識によってそれをさらに強化しています。

以前まで、Komodo社は、何千ものサイロ化したソースに散在していたデータを結び付けるのに苦労していました。Komodo社のITチームは、アクセス制御を構築したり、品質管理プロセスを維持したり、毎日発生する数テラバイトのデータを取り込めるよう拡張したりするのに悪戦苦闘していました。現在、Komodo Health社は、データクラウドを活用することで、150以上の支払者データセットとそれに対応する650億件以上の臨床および薬局データを組み合わせ活用しています。

データクラウドによってKomodo社の顧客は、RなどのフレームワークやPythonなどのプログラミング言語を使ってこのようなデータ資産を直接照会することができます。また、Komodo社は、顧客が自社のデータに対してカスタム分析を実行したり、独自のインサイトを獲得するために活用できるデータ資産を開発できるようにするデータサービスを提供しています。

## 小売業サプライチェーンを強化する

新型コロナウイルスの大流行が続く中、小売業者と消費財（CPG）メーカーは、スピード、利便性、品質を確保しながら、新しい配送チャンネルを通じて顧客のニーズを満たさなければならないという大きなプレッシャーにさらされています。小売業界の企業は、顧客需要の予測、サプライチェーンの管理、効率的なビジネス運営のためにデータをどのように活用すればよいかを再検討しています。小売業者は、仕入れる商品、実施するプロモーション、採用する価格設定モデルなどの業務上の課題を決定する際にもはや直感や推測に頼ることはできません。

小売業者とそのCPGパートナーの間には、データの交換に基づく互恵的関係があります。それぞれの組織が他方にとって有益な情報をサイロに保管しています。たとえば、小売業者は、CPG企業が容易にアクセスできない詳細な販売時点（POS）データを保有しています。CPG企業はこのデータを分析し、他のソースと組み合わせることでカテゴリーおよびブランド管理に関

する提案を行いたいと考えています。適切な商品を適切な数量だけ生産するためには、顧客が何を購入しているかを正確に知る必要があります。

このようなニーズを満たすため、小売業者とCPG企業の双方がコネクテッドサプライチェーンを構築しています。コネクテッドサプライチェーンでは、インテリジェントな需要予測によって自動補充戦略が推進され、顧客の求める商品が適切な場所に、適切なタイミングで、そして適切な数量だけ常に提供されることが保証されます。

このような消費者レベルおよび商品レベルのインサイトを得るためには、管理されたライブデータの完全なソースが不可欠です。ただ、残念なことに、取引データはPOSアプリケーション内にサイロ化していることが多いため、マーケティングアプリケーション、サプライチェーン管理アプリケーション、在庫管理システム、CRMシステム等の他のデータソースとの統合が困難です。



ヒント

顧客に関する調査、在庫レベルのより効率的な管理、人気商品の需要予測などに取り組む際には、POSデータが出発点となります。

データクラウドによって、小売業者とCPG企業はPOSデータなどのデータを簡単に共有しながら、相互に有益な関係を築き、パートナーネットワークを拡大することができます。共有データは、商品をサプライヤーから物流センター経由で小売店やEコマース拠点に移動するという作業を同期するための鍵となります。

小売企業やCPG企業もまた分析にデータクラウドを活用しており、その際にはSnowflakeデータマーケットプレイスで利用可能なサードパーティデータを頻繁に利用しています。たとえば、Numerator社は、米国の100万以上の家計からの消費者購買情報を提供し、市場開拓戦略に役立たせています。データプロバイダーのAdverityは、Google、Facebook、Snapchatの広告データを提供し、企業が顧客の嗜好や購買傾向に関する理解を深める支援を行っています。このようなデータサービスは、小売業者やCPG企業にとってデータを収益化して新たな収益源を生み出す機会となることもあります（図4-1参照）。

言うまでもなく、GDPRやカリフォルニア州消費者プライバシー法（CCPA）などの厳格な消費者プライバシー法とサイバーセキュリティの脅威によって、徹底したデータガバナンスと高度なセキュリティ対策の必要性が高まっています。すべての小売業者が顧客を知り、そのニーズを予測したいと考えていますが、現行のデータプライバシー法に準拠していない企業がPIIデータの保護を怠った場合、多額の罰金が科される可能性があります。

# 小売サプライチェーンのための安全な共有データ

データクラウドにより、小売企業はデータを一元管理し、サプライチェーンのパートナーとライブデータを安全に共有しながら、価格および在庫戦略の最適化、利益率の向上、消費者プライバシーの確保に取り組むことができます。

- **小売業者**：サプライチェーン、在庫、POS、CRM、顧客ロイヤリティ、およびマーケティング分析システムに散在するサイロ化されたデータを一元化します。
- **CPG企業**：小売業者の取引データにアクセスし、それを他のデータと組み合わせて分析し、ブランド管理のためのインサイトとして小売業者に提供します。
- **サプライヤー**：小売業者のオムニチャネル販売データにアクセスして商品の数量を予測し、在庫切れや過剰在庫を最小限に抑えます。
- **カスタマーエクスペリエンス**：小売およびCPGビジネス全体のデータにアクセスできるようにすることで、シームレスでパーソナライズされた体験を顧客に提供します。

## 小売業向けデータクラウド

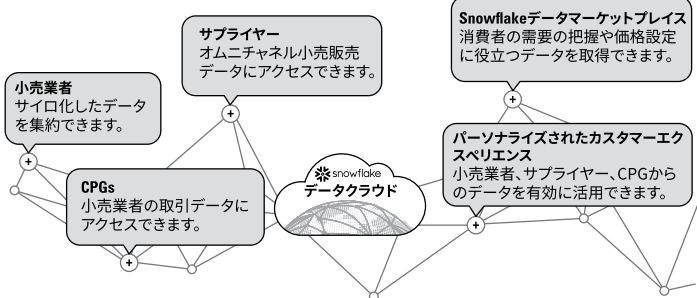


図4-1: データクラウドは、サプライチェーンのすべてのメンバーがデータを効率的に共有し、データから価値を引き出すインフラストラクチャに投資することを可能にします。

## 全国規模の小売業者がデータ範囲を拡大



ケーススタディ

Office Depot社には、ウェブサイト、モバイルアプリ、小売店、倉庫、サプライチェーン、車載IoTデバイスなど、さまざまなソースからデータが送られてきます。ビジネスアナリストは、財務実績の測定、企業のマーケティング戦略の強化、マーケットトレンドを明らかにするCustomer-360プロファイルの改善にこのデータを活用しています。

以前は、データウェアハウスが複数存在したため、アナリストは目的に応じてデータビューを使い分ける必要がありました。たとえば、あるタイプのリポジトリをウェブおよびモバイルデータに使用し、別のタイプのリポジトリをIoTデータに使用していました。データストレージ容量が限られていたため、どのデータをどのくらいの期間保管するか決定しなければならず、長期的な履歴分析を行う余地がありませんでした。現在、Office Depot社はデータクラウドを使用して必要な構造化データと半構造化データをすべて統合しています。ほぼ無制限のストレージとコンピュータ能力を手に入れたため、管理者は詳細な分析を以前よりもはるかに迅速に行うことができるようになりました。たとえば、ウェブログを分析し、同社のEコマースサイトでの消費者行動を明らかにするプログラムは、以前は実行に何時間もかかっていましたが、現在は数分で実行できます。さらに、Office Depot社は、過去の履歴データを、その価値がすぐには明らかにならない場合でも、無期限に保存しておくことができるようになりました。ストレージの容量不足のためにデータを削除する必要はもうありません。

また、Office Depot社はSnowflakeのセキュアデータシェアリングテクノロジーを活用して社内の一元化された場所にデータを保持しながら、ほぼ瞬時にアクセスできるようにして事業部門全体で共有しています。ビジネスチームは、社外と簡単にデータを共有できるほか、新しい機会が生まれたときにSnowflakeデータマーケットプレイス内の管理されたデータを収益化することも検討しています。



ポイント

組織は、データクラウドを採用することで、ますます厳格化するデータガバナンス要件への準拠を可能にするコントロールを導入することができます。

# 優れたメディアおよびエンターテインメントサービスを提供する

メディア産業は変化の真ただ中におり、従来のビジネスモデルはデジタルコンテンツとオンラインサービス（現在のテレビ市場を変革しているビデオオンデマンド（VOD）サービスなど）に対応できるよう進化しています。従来の放送局やケーブルテレビ会社は、Netflix、Huluをはじめとする数十社のコンテンツプロバイダーにその地位を揺るがされており、長年視聴してきた加入者の多くは、従来のケーブルパッケージ、衛星放送、セット販売方式のテレビコンテンツを「断捨離」する傾向にあります。既存のメディアプロバイダーは、VODプラットフォームの開発を急いでおり、市場には何百もの視聴方法が存在しています。テレビ局、ケーブルテレビ会社、ストリーミングコンテンツプロバイダー、制作スタジオ、そして支配的なテクノロジー大手企業であるApple、Google、Amazonなどが、この混沌とした部門で加入者をめぐる競争を繰り広げています。

他のタイプのメディア企業やエンターテインメント企業も、デジタルサービスの導入競争において同様の障害に直面しています。出版者や通信事業者に加えて、この活気ある産業には次のような企業も名を連ねています。

- ▶ **アドテック（Ad tech）企業**：マーケティング代理店とブランドがデジタル広告を展開し、その効果を分析するのを支援しています。
- ▶ **広告代理店、データ分析測定会社、データエンリッチメント会社**：メディア企業の顧客プロファイルの作成と拡充を支援するなど、クリエイティブなマーケティングサービスを提供しています。
- ▶ **アート展示施設およびスポーツ関連ビジネス**：ライブやオンラインでイベントやエンターテインメントを提供しています。
- ▶ **巨大なビデオゲーム産業**：パッケージ型、ネットワーク型、クラウド型のゲームを提供しています。

1つ変わっていないことがあります。それは、どんなメディア企業も、サブスクリプション、追加購入、広告収入から利益を上げるためには、オーディエンスを獲得し、維持し、引き付ける必要があるということです。データは、適切なコンテンツを決定し、顧客を獲得し、解約を防ぎ、視聴者のエクスペリエンスを最適化するための鍵となります。



ポイント

デジタルビジネスモデルは、顧客の絞り込み、パーソナライズされたコンテンツ提案、マーケティングおよびアウトリーチキャンペーンの効果測定に役立つデータを安定的に提供します。

残念なことに、多くの広告主やブランドは、必要なデータへのアクセスが制限されており、そうしたデータと他のソースを統合する方法を十分に理解していません。従来のメディア企業やエンターテインメント企業は、顧客データをさまざまサイロに散在させる古いテクノロジーを使用しています。このようなデータが分断された状況では、重要なビジネス目標と規制コンプライアンスの達成が難しくなります。データプライバシー法が厳格化されるに伴い、コストとリスクはますます大きくなると予想されます。

多くのメディア企業とエンターテインメント企業は、こうした欠陥を補うために最新のデータシェアリングを導入する必要があることを認識しています。データクラウドは、複数のプラットフォームやチャンネルに分散する消費者や加入者のデータを統合し、それに購買情報を結び付けます。データシェアリングは、利用者の個人情報を統一し、そのオンライン行動をすべて把握し、最終的に各加入者の顧客生涯価値（CLV）を明らかにするための出発点となります。

メディア企業とエンターテインメント企業は、このような完全な顧客ビューを作成した後、データクラウドを使用してこれらのインサイトへのアクセスを民主化し、チームがコンテンツの制作や獲得、配信、広告、製品体験について、より豊富な情報に基づいて意思決定を行えるようにすることができます。さらに、メディア企業は、オーディエンスに関するより詳細なインサイトをブランド広告主と共有することで、広告主が広告費用対効果をより適切に予測し、広告在庫の価値を高められるようにすることができます。

たとえば、広告主は、さまざまなソースからきめ細かい顧客データを収集して、より良い広告を作成し、広告キャンペーンの効果を測定します。メディア企業やエンターテインメント企業は、データクラウドを利用することで、グローバルに利用可能なプラットフォームにあらゆるデータを低コストで保管し、管理されたデータを事業部門全体で、広告主と、そしてデータエンリッチメントパートナー間で共有できます。



ポイント

データクラウド内のSnowflakeデータクリーンルームでは、データを移動したり、PIIを公開したりすることなく、メディア企業、広告主、その他のパートナー間でファーストパーティデータを安全に共有できます。企業は、Snowflakeを介してデータを共有することで、価値ある消費者データを取得・交換するときにデータプライバシーを管理・コントロールできます。

さらに、Snowflakeデータマーケットプレイスには関連するデータセットとデータサービスが用意されており、商用データプロバイダーからのエンリッチメントサービスと組み合わせて使用することで、アイデンティティグラフ、デモグラフィックデータ、オーディエンスデータを簡単に取得できます（図4-2参照）。

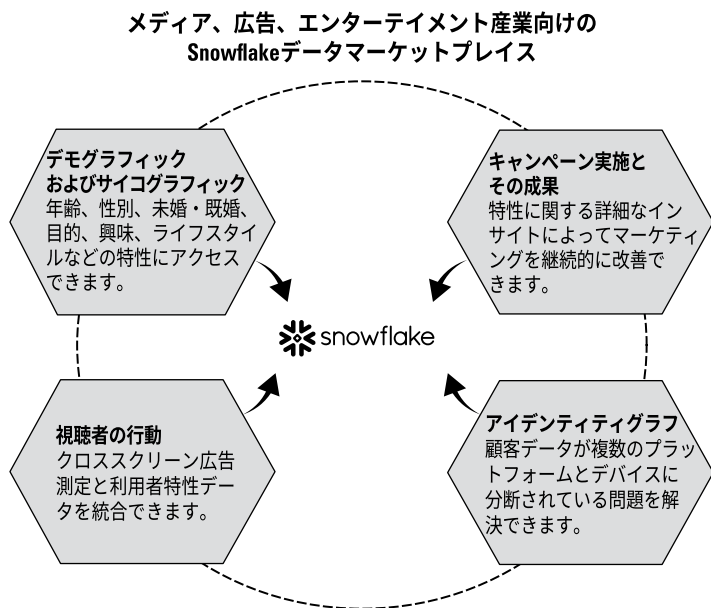


図4-2：Snowflakeデータマーケットプレイスは、メディアや広告を含むさまざまな業界向けのサードパーティデータとデータサービスを調達するためのネットワークです。

## メディアおよびエンターテインメント向けの優れたエクスペリエンス

メディア企業はデータクラウドを使用して、ブランド、広告主、プラットフォーム、エンリッチメントサービスプロバイダー間で管理されたデータを安全に共有できます。その結果、これらの組織は、加入者の生涯価値と広告費用対効果を高めるための意思決定をデータに基づいて下すことができます。

- **メディア企業**：複数のプラットフォームとチャンネルにまたがる加入者の行動データをブランドからの購入情報と統合し、360度の顧客ビューを作成します。

- **ブランド**：複数のメディアチャネルと広告プラットフォームにわたって顧客接点を統合し、コンバージョンと売上を増やすためのキャンペーンを最適化します。
- **広告主**：広告費とメディア消費および購買行動を結び付けて、マーケティングプログラムの投資利益率を明らかにします。
- **データプロバイダー**：データクラウドを使用して自社データと顧客のデータをシームレスに結びつけることで、デジタルおよびオンラインサービスの規模を拡大します。

## 公共サービスの質を向上させる

政府機関は、データクラウドを活用することで、政府の枠を超えるデータフローを最新化できます。たとえば、政府機関全体での収集分析を可能にして科学研究を促進したり、連邦政府機関と州政府機関の間のデータコラボレーションを強化したり、市民と外部利用者へのライブの安全なデータ配信を改善して公益に寄与したりできます。

クラウドベースのテクノロジーを導入する動きは政府のあらゆるレベルでITインフラに変革をもたらしています。連邦政府、州政府、および地方自治体は、データの収集方法、保存方法、分析方法を最新化するための大規模な施策を実施しています。これらの施策はいずれも、政府機関がその使命を果たし、米国行政予算局の年次戦略行動計画などの戦略的なIT指令を履行するためのものです。さらには、政府機関にとっては、市民サービスの向上、行政および業務の非効率性の解消と合理化、研究やイノベーションの推進、機関間のコラボレーションのサポートのためにデータを活用する機会や必要性がかつてないほど高まっています。

データクラウドは、シームレスかつ安全なデータ共有を通して、政府機関が機関内、部門間、異なる政府レベル間、そしてクラウド間のデータサイロを効果的に解消できるようにします。同じく重要なのは、データクラウドがすべての市民データへの安全で管理されたアクセスを確保できることです。



ポイント

Snowflakeは、データクラウドが政府組織に求められるプライバシーおよびセキュリティ規制を満たすという困難な仕事を成し遂げ、FedRAMP、SOC、FISMA、NIST、FIPS規格を含む規制への準拠を継続的に監視しています。

政府機関では、新型コロナウイルスの発生により、従来のデータシェアリングに取って代わる最新の方法を見つけることがますます急務となっています。カリフォルニア州は、データクラウドを活用して郡の職員、さまざまな州政府機関、住民に州全体にわたるシームレスなデータアクセスを提供し、最新データがそれを最も必要としている人々の手に届くようにすることで、調和のとれた公衆衛生上の対応が取れるようにしています（詳細は第1章のケーススタディを参照）。同州の共有データセットには、カリフォルニア州公衆衛生局の症例データのほか、病院や研究所のデータが含まれています。さらに、カリフォルニア州はSnowflakeデータマーケットプレイスを活用して、州全体の症例、検査、人口統計、ホームレスへの影響などの最新情報を一般に公開しています。

言うまでもなく、公的部門は、コストを抑えながら効率性を高め、市民や企業へのサービスを向上させる施策を常に模索しなければなりません。公的部門の組織は、データをデータクラウドに置き、Snowflakeデータマーケットプレイスで成長し続けるサードパーティデータサービスを利用することで、より効果的で経済的な情報システムを構築することができます。政府機関は、あらゆる種類のデータに安全かつ自由にアクセスするという喫緊のニーズに応えるためにデータクラウドを活用することで、重要な情報を安全に共有するためのシンプルで費用対効果の高いソリューションを手に入れることができます。

## 政府機関のコラボレーションによる市民サービスの向上

公的部門の組織は、データクラウドを活用して、機関および部門のガバナンスポリシーを遵守しながら、重要なデータへのライブの取り消し可能なアクセスによってコラボレーションを強化しています。

- **連邦政府機関**：運営データと管理データに安全にアクセスし、政府機関、州政府、市民と共有します。
- **州政府/地方自治体**：政府機関、市民、パートナーからライブの安全なデータを受け取って共有しながら、データ管理とガバナンスを向上させます。
- **エコシステムパートナー**：政府は、公的部門と民間部門のパートナー間のデータシェアリングに伴うリスクを軽減し、その効率性を高めることで、部門の枠を超えた関係を強化できます。
- **市民**：安全でほぼリアルタイムのデータアクセスによる、より効果的で効率的な行政サービスを市民に提供します。

- » 適切なアーキテクチャを使用して開始する
- » データクラウドを歴史的な視点から考える
- » ほぼ無限の拡張性とパフォーマンスを解き放つ
- » データとワークロードを複数のクラウドに拡張する
- » 強力なセキュリティとガバナンスを適用する

## 第5章

# Snowflakeのプラットフォームを掘り下げる

Snowflakeの創業者が目指したのは、それまでの30年間のコンピューティングに破壊的な影響を与えることでした。データ管理とデータ運用のあらゆる側面を検討し、無限に近いパワーと拡張性、そしてワールドクラスのデータガバナンスとセキュリティを実現するプラットフォームの構築を目指しました。その結果として誕生した、クラウドを基盤とした新しいプラットフォームがデータクラウドの礎となったのです。本章ではこのプラットフォームを詳しく解説していきます。

## 適切なアーキテクチャを使用して開始する

Snowflakeのプラットフォームは何よりもまずクラウド向けに構築されたもので、従来古い技術に制約を受けることはありません。パブリッククラウドサービスの可用性、受容性、実用性の持続的な向上を前提とするSnowflakeは、複数のパブリッククラウドやリージョンに展開することができる一方で、単一のインターフェースを介してもアクセス可能です。多様な分析ワークロードを大量に処理することが可能で、ほぼ無制限の量と種類のデータを最小限のレイテンシーで処理することができます。

# クラウドで構築された独自のアーキテクチャ

プラットフォーム独自の機能は、次の4つの画期的なテクノロジーに支えられています。

- ほぼ無制限の拡張性、同時性、効率性を実現する**マルチクラスター共有データアーキテクチャ**
- 複数の**パブリッククラウド**を1つのクラウドのように扱うことのできる**統一されたコードベース**
- 卓越したデータ保護を実現する**組込み型のデータガバナンス、セキュリティ、機能**
- データの移動やコピーを必要とせず、複数の組織が相互にライブデータを瞬時に共有・受信できるセキュアな**データシェアリングテクノロジー**

この独自のアーキテクチャにより、組織内はもとより、事業パートナー、顧客、急速な成長を遂げるSnowflakeデータマーケットプレイス、データクラウドを構成する何千もの組織など、さまざまなタイプのユーザーによるデータを使った自由なコラボレーションが実現します。

## 過去の教訓を活かす

従来のデータプラットフォームは限られたコンピューティングリソースを活用するように設計されており、大抵はオンプレミスのデータセンター内に設置されていました。新しいデータウェアハウス、データレイク、データマート、その他のワークロードのサイズを決めるには、慎重なキャパシティプランニングが必要です。こうしたデータ環境がどれほど普及するかは事前に把握できないため、組織はプロビジョニングを過剰にせざるをえません。当初必要とされるよりも多くのハードウェア、ソフトウェア容量を導入しなければならないのです。さらに、データドリブン型ワークロードの多くは、会計士が月末に決算を行うときや、小売業者がショッピングのピーク時に結果を毎年予測したり分析したりしなければならない場合のように、活動が活発になる時期がまちまちに来ることが特徴です。こうした短い期間にだけ容量のニーズが発生するため、ピーク時を前提として環境のサイズを決定するのは無駄が多いのです。

分析アプリケーション、データ・サイエンスアプリケーション、データエンジニアリングパイプライン、その他多くの種類のデータアプリケーションの普及と重要性の高まりによって、対応しきれない従来のプラットフォームは時代遅れになりました。



技術情報

従来のデータプラットフォームは拡張性に乏しく、コンピュートやストレージのリソースが固定されているため、複数のユーザーが同じデータやコンピュートリソースに同時にアクセスできる尺度であるコンカレンシーが制限されます。アーキテクチャは直線的なため、並行して複数のワークロードを実行できず、リソースやデータに基づくインサイトを得られるまで長く待たなくてはなりません。

パブリッククラウドサービスの台頭により、企業は突然、無限の計算能力とストレージ容量を利用できるようになりました。これにより、理論的には、従来のデータ環境でもより多くのユーザーやワークロードをサポートできるようになりました。しかし、従来のデータベース管理システムや分析プラットフォームは、このパワーや容量を活用できるように設計されていません。多くのシステムは単にクラウドに「移動&移行」しただけで従来の制約の下で稼働し続けていました。これらの情報システムは、限られたリソースで動作し、1つの種類のデータを利用するように設計されています。構造化データ用のデータウェアハウス、半構造化データタイプ用のデータレイク、そして多種多様なローカルデータベース、データマート、運用データストア（クラウド上のものもあれば、オンプレミスのももある）など、それぞれが自部門の独自のニーズを満たすよう構成されています。

## パフォーマンスを高めコストを削減する

Snowflakeのプラットフォームは、クラウドの特性を最大限に活用できるように設計されています。コンピュートとストレージのリソースはそれぞれ、各アプリケーション特有のニーズに対応して独立して拡張可能です。たとえば、データエンジニアリングパイプラインを介してデータを取り込みつつ、同時に同じデータを使用して機械学習モデルをトレーニングするなど、ほぼ無制限のユーザーがパフォーマンスを低下させることなく、同時にワークロードを実行することができます。

Snowflakeのプラットフォームでは、データが保存される中央リポジトリ、独立したほぼ無限のコンピュートリソース、セキュリティ、ID管理、トランザクション管理などのクラウドサービスのレイヤー、その他の多くの機能など、すべての部分が一体となって存在します。

## 適切なアーキテクチャが重要な理由を理解する

なぜ技術的なブレークスルーがそれほどまでに重要なのか、そして、企業にはどんなメリットがあるのかを考えてみましょう。Snowflakeのお客様は、新しい多様なワークロードを制限なく立ち上げ、必要なだけの計算能力とストレージ容量を秒単位でプロビジョニングすることができます。ストレージと計算リソースを別々に拡張できることに加えて、使用したリソースにだけ料金を支払うこともできます。また、同じデータにアクセスするワークロードであっても、各ワークロードには専用のコンピュータクラスターが割り当てられます（図5-1参照）。

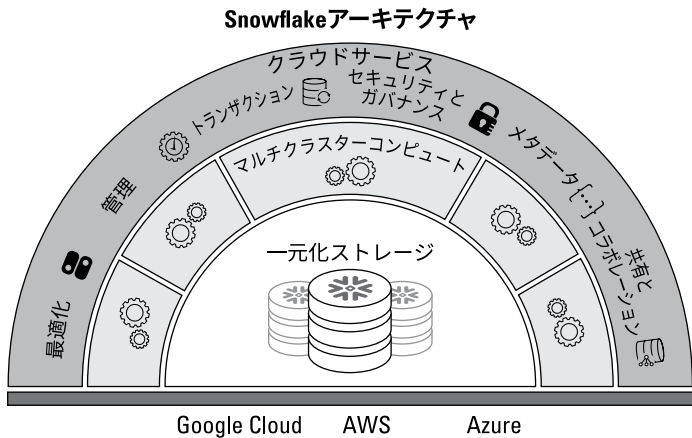


図5-1：Snowflakeのプラットフォームでは、ストレージをコンピューティングリソースから分離することで、複数の部門（さらには複数の組織）が同じデータに対してほぼ無制限の数のワークロードを同時に実行できます。



ポイント

Snowflakeのプラットフォームが実質的に無制限の拡張性と弾力性を提供できるのは、各ワークロードの規模に応じて動的に展開できるアクティブなクラスターの数に実質的な制限がないためです。いつでも必要なだけリソースを確保でき、アイドル状態の容量に料金を支払う必要はありません。



ポイント

特許取得済みのマルチクラスター、共有データアーキテクチャを基礎に構築されたこのプラットフォームでは、パフォーマンスに影響を与えることなくほぼ無制限での同時実行が可能です。計算リソースはその場で、または自動的に拡張されるため、その瞬間に発行されるクエリの数にかかわらず、ユーザーは安定したパフォーマンスを享受できます。

## パフォーマンスと拡張性を解き放つ

Snowflakeのプラットフォームは、次のようなパフォーマンス強化サービスを備えています。

- **マイクロパーティショニングおよびクラスタリング機能**は、クエリの実行中にデータを動的に削除します。これにより、Snowflakeはクエリに関連するデータのサブセットのみをスキャンします。
- **マテリアライズドビュー**（クエリの結果を含むデータベースオブジェクト）は、一般的な繰り返しの多いクエリパターンのパフォーマンスを向上させ、リソースの多いクエリを高速処理します。参照先のベーステーブルにデータが追加・変更されると、Snowflakeはバックグラウンドで自動的かつ透過的にマテリアライズドビューを更新します。
- **クエリアクセラレーションサービス**は、追加の計算リソースや優先順位からメリットを受けることができるクエリ部分を自動的に特定します。クエリアクセラレーションは、データサイエンスやスキャン多用型のワークロード処理に特に有用です。
- **検索最適化サービス**は、クエリをさらに高速化します。たとえば、文字列内のパターンマッチングを含む検索を行うことで、ログデータの分析などのユースケースが増え、サイバーセキュリティ分析のパフォーマンスが向上します。

## マルチリージョン、マルチクラウドを1つのサービスで

データクラウドは、主要なパブリッククラウドと、その多くが位置する4大大陸のリージョンに対応しています（図5-2参照）。Snowflakeのリージョンは、Amazon Web Services、Microsoft Azure、Google Cloud Platformといった主要なクラウドプロバイダーが提供するインフラ上に構築されています。

データクラウドでは、共有されたメタデータレイヤーが集約された一連のネットワークサービスを定義し、リージョンやクラウド間でのデータやワークロードの転送を容易にしています。すべてのデータクラウドの分析ワークロードは一貫して実行され、すべてのユーザーは、使用されているリージョンやクラウドプロバイダーを問わず、1つの一貫性あるエクスペリエンスを体感することができます。Snowflakeは、世界のどこであっても、クラウド上のすべてのデータのトランザクションの一貫性を常に確保します。データベースの操作が実行中に中断された場合でも、データベースに変更を加えることなく、データはそれ以前の状態に「ロールバック」されます。

## データクラウド クロスリージョンとクロスクラウド

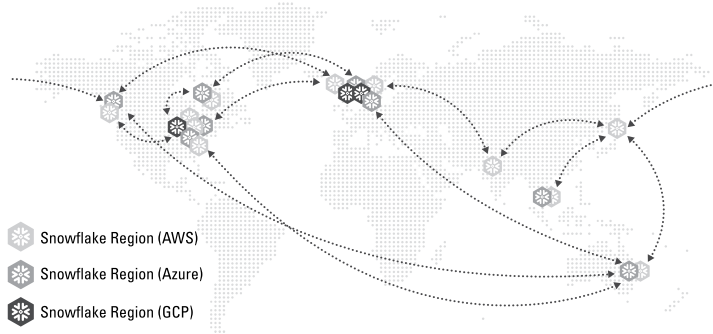


図5-2: データクラウドは、4大陸と3つの主要クラウドプロバイダーのインフラストラクチャをカバーしています。

## 使いやすいプラットフォームを享受する

本書を通して説明しているように、データクラウドでは、インフラストラクチャをプロビジョニングしたり、複雑なソフトウェア環境を管理したりすることなく、必要なデータアプリケーションを開発して展開できます。ほとんどの場合、プラットフォーム自体が管理を担っています。パフォーマンスの調整や最適化は必要ありません。データベース管理者（DBA）は、インデックス管理やデータをパーティショニングする必要はなく、プラットフォームが自動的に各ワークロードを管理し、パフォーマンスを最大化します。DBAはプラットフォームの管理から解放され、データから価値を引き出すことに注力できます。

このプラットフォームは、データの取り込み、データアプリケーションの開発、管理されたデータのエコシステム全体での安全な共有など、さまざまなプロセスの中心になります。課金は秒単位で行われるため、ユーザーや作業グループは、使用した正確なストレージおよびコンピュートリソースについて料金を支払うことができます。

### 使用量の予測と監視

リソースの監視・管理機能が組み込まれているため、使用状況や課金状況を完全に把握し、個々の予算に応じてチャージバック機能とショーバック機能をきめ細かく設定できます。Snowflakeは、以下の2つの方法で、お客様による全体的な使用状況の監視、管理、予測を支援します。

- ▶▶ 監視コントロールでは、プラットフォームが予測通りにクレジットを消費しているかどうかを監視します。
- ▶▶ 管理コントロールでは、使用量を予測可能にします。

お客様は常にコントロールを維持できるだけでなく、使用量の急増を防止するセーフガードを講じることもできます。たとえば、自動停止・自動再開機能は、Snowflakeのデータ処理が一時中断される場合、リソースアカウンティングを自動的に開始および停止します。また、使用する顧客がワークロードの種類ごとに特定のタイムアウト期間を設定することもできます。たとえば、ETL（抽出、変換、読み込み）ジョブの場合は「完了後すぐ」、ウォームキャッシュを必要とするクエリ操作の場合は「10分」のように設定できます。

## あらゆるタイプのデータへ容易にアクセス

データクラウドでは、標準的なSQLを使用して、構造化されたデータや半構造化されたデータの保存、管理、問い合わせを行うことができます。Snowflakeは今後のリリースでオーディオ、ビデオ、PDFなどの特定のタイプの非構造化データをサポートする予定です。

Snowflakeにデータを保存する場合でも、クラウドインフラストラクチャプロバイダーのオブジェクトストアなどの外部リポジトリに保存する場合でも、すべてのユーザーは一貫性のあるインターフェイスを使用してデータを照会、表示、管理できます。ユーザーは、メインプラットフォームからデータにアクセスする場合と同様に、外部テーブルのデータに簡単にアクセスできます。共通のメタデータレイヤーにより、すべてのユーザーが一貫した結果を得られ、すべてのワークロードが一貫した結果を生み出します。

データクラウドは、さまざまな種類のデータやデータワークロードに共通のリポジトリを確立することで、データ管理タスクを簡素化し、サイロを解消します。データが一元化されることで、ユーザーがデータを使用するまでに必要な段階を減らすことができます。データの時間的価値が高まることで、ユーザーはより迅速に行動し、業務上のビジネスプロセスにアナリティクスを導入できます。



ポイント

Snowflakeのプラットフォームにはメンテナンスがほとんど不要で、組織は、業界や地域のデータプライバシー要件を満たしながらも、異なるパブリッククラウドやリージョンでも単一のクラウドとして運用することができます。

# 強力なセキュリティとガバナンスを適用する

複数のクラウドプラットフォームにまたがって統一されたコードベースを使用することは、強力なデータセキュリティを適用する上でも重要です。Snowflakeは、すべてのパブリッククラウドプロバイダーに対して、同様のセキュリティ設定と管理手法を適用しています。サイバーセキュリティの担当者は、監査証跡やイベントログの差異を解決したり、データを暗号化するために複数の管理システムを操作したりする必要はありません。暗号化はデフォルトで行われ、データはお客様が管理する暗号化キーに従って転送中も保管中も暗号化されます。



ポイント

Snowflakeのサイバーセキュリティチームは、包括的なモニタリングを行い、常時発生するアラートを分析し、業界をリードするサイバーセキュリティプラクティスを実施して脅威を阻止し、問題を解決します。

ロールベースのアクセス制御によって、ユーザーは明示的に閲覧を許可されたデータにのみアクセスできます。こうした制御は、テーブル、スキーマ、外部テーブルなどの仮想拡張機能を含むすべてのデータベースオブジェクトに適用され、複数の種類のデータ、異なる多くのリージョン、複数種類のパブリッククラウドが含まれる場合でも適用されます。Snowflakeは、共通のソフトウェアサービスレイヤーを介すことでパブリッククラウドによる違いを解消するため、パブリッククラウドごとに専用のスキルセットを持つ担当者を置く必要はありません。

ダイナミックデータマスキングや行アクセスポリシー機能などのポリシーベースのセキュリティ制御は、データを照会する各ユーザーに与えられた権限に基づいて値を編集したり、ユーザーの役割に応じてデータの可視性を制限したりします。

セキュアビューは、給与や社会保障を含むフィールドを隠したり、顧客の個人識別情報（PII）を保護するなど、アクセスをさらに制限することができます。新しいタグ付けフレームワークにより、ユーザーはオブジェクトにメタデータを添付し、これらのタグに基づいてポリシーを適用することができます。

最後に、このプラットフォームでは、データへのアクセスルールを定義する行レベルのアクセスポリシーを簡単に設定することができます。Snowflakeのクエリエンジンは、ユーザーに対応するルールを動的に適用し、各ユーザーが閲覧可能な行のみを表示します。標準SQL構文を使用してこれらのアクセスポリシーを実装することで、データベース管理者は、ポリシーを簡単に作成・更新することができます。

## 機械学習やアナリティクスのための連続データパイプライン



ケーススタディ

Convoy社は、機械学習と自動化ソフトウェアを使用して荷主と輸送業者の連携を図り、毎年何百万ものトラック運搬貨物を効率的に運ぶデジタル貨物ネットワークです。Convoy社のデータソリューションが基盤としているのは、データウェアハウスと、新製品の機能が重要なビジネス指標にどのような影響を与えるかを分析する、データサイエンティストによる実験サービスです。これは、Convoyの機械学習プラットフォームの基盤でもあり、より正確なマッチング、価格設定、より良いサービスを顧客に提供しています。

ただし、Convoy社の既存のソリューションは非効率でした。アナリストは、数時間ごとにApache Airflow経由で一括して読み込まれるデータに依存していました。これでは、アナリストが最新のデータにアクセスできなかっただけでなく、エンジニアが分析のためにデータを本番データベースに長期間保存する必要があったため、非常に高額なストレージコストがかかっていました。Convoy社は、データの取り込み、機械学習、アナリティクス、データサイエンスなど、複数のワークロードを競合させずに同時実行できる、より頑丈なデータプラットフォームを必要としていました。ソリューションはJSONおよびVARIANTデータタイプに対応し、データアプリケーションの構築、データパイプラインの設計、ビジネスインテリジェンスアプリケーションの作成、権限付与のための、Convoy社が確立した広範なツールのエコシステムをサポートできるものでなければなりませんでした。

そこでConvoy社が選択したのはSnowflakeです。マイクロバッチ処理による収集サービスでデータレイテンシーを削減し、ワークロードのパフォーマンスを最大化するためストレージリソースとコンピュートリソースを独立して拡張できるようにし、ネイティブフォーマットの半構造化データのサポートも実現しました。現在、Convoy社は、数時間ごとにApache Airflowジョブを実行する代わりに、SnowpipeとSnowflake Kafkaコネクタを使用して数千のマイクロトランザクションをわずか数秒で実行し、Amazon S3からSnowflakeのデータクラウドにデータを取り込んでいます。この連続データパイプラインにより、エンジニアは本番のデータベースにデータを長期間保存する必要がなくなり、ストレージの要件を軽減できました。Snowflake Streams and Tasksは、トランザクションが発生してから数分以内で新規データをデータクラウドに読み込みます。すべてのワークロードのパフォーマンスが大幅に向上しました。Convoy社は、このSnowflakeソリューションの導入によって、データの取り込みにかかる時間を20分の1に短縮し、クエリの応答を10倍高速化しました。



ポイント

データクラウドにデータを置くことでデータガバナンスが簡素化されます。ビジネスユーザーにとっては、業界固有・地域固有のデータ主権要件に準拠することが容易になり、企業の機密データや個人を特定できる情報（PII）が保護されます。Snowflakeでは、あらゆるデータ、あらゆるワークロードに対してガバナンスとセキュリティのプラクティスが統一されているため、外部テーブルに保存されたデータにアクセスする場合でも、適切なコントロールを維持できます。共通のプラクティスの下で適用されるデータアクセス制御は、拡張したエコシステム全体の使用状況の監視と監査に役立ちます。データが異なるクラウドに存在する場合でも、多くのワークロード、関連するすべてのSnowflakeアカウントに対して一貫したポリシーを適用できます。

## 本章の内容

- » データウェアハウスとデータレイクの価値を最大限に高める
- » データサイエンスの取り組みを簡素化する
- » 頑丈なデータパイプラインを設計する
- » 最新のデータアプリケーションを作成する
- » データを制限なしに共有する

# 第6章 あらゆるワークロードを 実行する

**S**nowflakeのプラットフォームは、管理されたデータをソースから利用者に移動するエンドツーエンドのデータパイプラインを基盤に最も多様で重要な分析データのワークロードをサポートするように設計されています。本章では、顧客がデータクラウドに導入する最も一般的なワークロードである、データウェアハウス、データレイク、データエンジニアリング、データサイエンス、データアプリケーション、データシェアリングについて解説します。

## データウェアハウスとデータレイクの導入

データウェアハウスは、主にリレーショナルデータベースの構造化データを保存・分析するために設計されました。データレイクは、大量の生の半構造化データをネイティブ形式で保存するために生まれました。初期のデータレイクは、オンプレミスのHadoopデータベースを使用して構築されていました。最近では、Amazon Simple Storage Service (S3)、Microsoft Azure Blob、Google Cloud Storageなどのクラウドベースのオブジェクトストアがデータレイクの構築に使用されるようになりました。

こうした最新のクラウド型データレイクソリューションにより、IT担当者はハードウェアスタックの管理作業から解放されましたが、ソフトウェア環境の構築、統合、管理は引き続き必要です。これには、データを変換するための手順の設定、ユーザーを識別するためのポリシーや手順の策定、データの暗号化、効果的なデータガバナンスの実施など、時間を要する多くの作業が必要です。これらをすべて終えてからでないとう有用な分析の作成という重要な作業にとりかかれません。

対照的に、データクラウドは、世界で最も広く使用されている分析ツールやデータ可視化ツールをサポートする業界標準のSQL言語を使用して、あらゆるタイプのユーザーが卓越したパフォーマンスでデータを管理できるようにします。エンタープライズデータウェアハウス、最新のデータレイク、クラウド機能の長所を兼ね備えており、あらゆるデータやワークロードに対応できます。構造化データ、半構造化データ、および特定のタイプの非構造化データ（Snowflakeが今後のリリースで提供する予定の機能）を、データクラウド自体またはオブジェクトストレージサービス内の外部テーブルに生の形式でステージングできます。

## コアワークロードの強化

データウェアハウスは、組織がデータ分析を最新化するためにSnowflakeを活用した最初のワークロードでした。組織はSnowflakeを使い続けながら、データクラウドにワークロードを追加していきました。Snowflakeは、こうしたコアワークロードに対応できるように継続的に機能を強化しています。たとえば、データクラウドは現在、GEOGRAPHYデータタイプを使って地理空間データをサポートしています。GEOGRAPHYデータタイプは、球面大地の座標を使用して地球表面の点、線、多角形などの特性を保存します。ユーザーはいくつかの対応フォーマットからデータを読み込むことができ、Snowflakeはクエリを高速に実行できるように最適化された表現を保存します。

また、Snowflakeは、第5章で触れた検索の最適化機能やクエリの高速度機能など、データアクセスを向上させる新機能も定期的に追加しています。

データクラウドは、データレイクの価値を最大限に高める新機能も備えています。この機能は特に重要です。ウェブログ、モノのインターネット（IoT）デバイス、ソーシャルメディアネットワークなどのソースから得られる半構造化データを活用することに価値を見出している顧客がますます増えており、このようなデータの多くがこれまでデータレイクに保存されてきたからです。今では、顧客はこのデータをデータクラウド自体に保存することも、オブジェクトストアに保存して外部テーブル経

由でアクセスすることもできます。顧客は、データクラウドが提供するパワー、セキュリティ、そしてガバナンスをデータプール全体で活用できます。

また、顧客がハイブリッドクラウド戦略を追及するためのコントロールを最大限に高める新機能もあります。たとえば、データクラウドを使用してデータを変換し、Snowflakeのデータエクスポート機能を使ってそのデータをデータクラウドからフラットファイルや外部テーブルにプッシュすることができます。Snowflakeは、このような定期的なロード・アンロードプロセスを可能にすることで柔軟性を最大限に高め、1つの一貫したデータクラウド戦略内で多数のストレージオプションを組み合わせることができる柔軟性を顧客に提供します。

## その他の重要なワークロードをデータクラウドで実行する

多くの顧客は、データクラウドをデータウェアハウスやデータレイクの基盤として使用することに加え、Snowflakeでデータエンジニアリング、データサイエンス、データアプリケーションワークロードを実行しています。その理由は、データクラウドは単にデータを保存するだけのソリューションではないことです。データを処理するための実証済みのソリューションでもあり、データパイプライン内で複雑な抽出、読み込み、変換（ELT）操作を実行したり、データサイエンスワークロード内で機能エンジニアリングを実行したり、現在のデータアプリケーションの厳しいコンピュータ要件に対応したりできます。データクラウドにデータを置いたら、そこでデータを処理するのが理にかなっています。

### データパイプラインのエンジニアリング

データエンジニアリングでは、分析のためにデータを収集、準備、変換、配信します。これは、多くの場合、データの移動を自動化し、特定のタイプの分析に備えてデータを特定の形式に変換するデータパイプラインを使用して行われます。データクラウドは、データフォーマットの操作、データシステムの拡張、データ品質とセキュリティの確保などを行ってデータを使用可能な形にすることで、データを「本番対応」にするプロセスを簡素化します。また、データクラウドは、データエンジニアがデータパイプラインを作成してデータの取り込みおよび変換タスクを簡素化するのにも役立ちます。



ヒント

SnowflakeのSnowpipeサービスでバッチやマイクロバッチのアップロードを管理し、ストリーミングデータに対応することで、ユーザーは新しいデータをすぐに活用することができます。たとえば、小売店の店頭で商品がスキャンされるとすぐにデータが生成され、在庫を補充するよう促されます。銀行は、不正の疑いがあるクレジットカード取引を検知するとすぐに、カード所有者に問い合わせを行い、その取引を承認するか拒否するかを確認します。

また、データクラウドにはSnowflake Kafkaコネクタを介してイベントストリームを接続するためのユーティリティも用意されています。開発者は、SQL、Java、Scala、Pythonを使用してコードを記述し、データクラウド内でデータフレームなどのおなじみの概念を使用してそのコードを実行することで、データを列や行に読み込むことができます。データエンジニアは、複数のストリームとタスクを調整する高度なデータ取り込みパイプラインをデータクラウド内で構築できます。SnowflakeのSnowparkサービスは、2020年11月現在、テスト環境で利用可能であり、データエンジニアは一般的なプログラミング言語を使用してデータパイプラインを開発し、データクラウドの内部または外部でデータを処理することができます。Snowflakeは、将来のリリースでこのサービスを一般公開する予定です。

## データサイエンスを簡素化する

データクラウドは、機械学習、人工知能、予測アプリケーションの開発からなるデータサイエンスのライフサイクル全体を合理化します。これにより、データサイエンティストは、新しいアプリケーションの開発に必要なデータを迅速に取得し、関連する機械学習（ML）モデルをトレーニングすることができます。データクラウドは、複雑なMLアルゴリズムを内部で実行し、Snowflakeの外部関数フレームワークを介して外部のMLサービスと統合することができます。たとえば、Java、Python、Scalaで記述されたリスクスコアリングモデルを呼び出し、結果セットを取り込むことができます。データサイエンティストやデータエンジニアは、これらの手順を外部で実行したり、データクラウド内で直接実行してプロセス全体を簡素化し、データクラウドをデータの準備や機能エンジニアリングに最適なものにすることができます。

## データアプリケーションの構築

これまで、顧客の30%以上が、ほぼ無制限の拡張性と費用対効果に優れた従量課金制モデルを理由に、Snowflakeをデータアプリケーションのバックエンドとして使用しています。データクラウドは、アプリケーションの開発やリソース集約型のワークロードの実行のために、ほぼ無限の計算リソースを提供します。これは、モバイルアプリケーションの開発者やSaaS（Software-as-a-Service）プロバイダーが、データの収集、処理、分析に重点を置いたビジネスを立ち上げる中で、ますます重視される機能となっています。

## データサイエンティストの選択肢

データサイエンティストがデータクラウドを選ぶ主な理由は2つあります。

**すべてのデータへのアクセス：**データサイエンティストは、MLモデルの構築とトレーニングのために膨大なデータを必要とします。データクラウドでは、事実上すべての関連データを1つの場所に保存することができ、最高レベルのガバナンスを維持しながら、さまざまなデータサイエンスノートブックやAutoMLツールにデータをインポートするプロセスを簡素化することができます。

**データプログラマビリティ：**Snowflakeは外部のMLサービスと統合しているため、Java、Python、Scalaなどのプログラミングフレームワークがサポートされるデータクラウドの内部で高度なアルゴリズムを直接実行できます。これによって、データサイエンティストやデータエンジニアは自分の好みの言語でプログラムを作成する柔軟性が得られます。

従来のソフトウェア企業は、データクラウドを活用して既存のアプリを最新仕様にするすることができます。こうした企業の多くは、クラウドが登場する前に作成された古いオンプレミスのクラウドウォッシュされたデータウェアハウスなど、従来のデータスタック上に製品を構築していました。このようなソリューションには、データクラウドを強力なものにするクラウドネイティブな特性がありません。従来のデータアプリの多くは、アプリケーションログ、Webサイト、モバイルデバイス、IoTセンサーなどから、一般にスキーマレスの半構造化形式で提供される膨大なデータを処理するのに苦労しています。



技術情報

データクラウドは、新しいデータアプリの開発、イテレーション、テスト、品質保証（QA）のために、ほぼ無制限のコンピュートリソースとストレージリソースを提供します。また、Python、Node.js、Go、.NET、Java、SQLなど、今日のソフトウェア開発者が使用するクラウドネイティブな開発ツールをサポートしています。

データクラウドは、スムーズなデプロイメントのためのデータリカバリ機能も備えています。たとえば、すべてのソフトウェアリリースとデータが適切にバックアップされるようにするため、Snowflakeは継続的なデータ保護サービスを提供しています。また、従来のバックアップスクリプトとプロセスを不要にする組み込みの機能も提供しています。

データクラウドは、主要なアプリケーション開発ツールやプログラミング言語に対応しているだけでなく、継続的インテグレーション（CI）や継続的デプロイメント（CD）などのDevOpsの原則を活用するように設計されており、開発者は、構造化データ、半構造化データ、および特定のタイプの非構造化データ（Snowflakeが将来のリソースで提供することを予定している機能）を標準的なSQLで照会できます。

データエンジニア、データサイエンティスト、そしてデータアプリケーション開発者は、コンピュート負荷やデータ負荷の高いワークロードをデータクラウドに移行することで、IT環境を劇的に簡素化することができます。MLモデルの構築から複雑なETL変換の処理まで、Snowflakeのマルチクラスター共有データアーキテクチャは、こうしたすべてのニーズに対応する頑丈な処理エンジンを提供し、その機能は年々向上しています。

## データを制限なしに共有する

本書を通して説明しているように、データクラウドは、コラボレーションやデータ共有を望んでいる組織に独自の機能を提供しています。社内のデータシェアリングはもちろん顧客間のビジネスパートナーシップの一環として、またはSnowflakeデータマーケットプレイスで公開されているデータとデータサービスにアクセスすることによって行われる、データクラウドの他の顧客との外部データシェアリングも促進します。データセット全体または選択したサブセットを他のSnowflakeアカウントと共有することができます。さまざまなチームが、データの複数のコピーを保持したり、データを移動させたりすることなくコラボレーションできます。一貫したデータガバナンスにより、誰がどのデータを閲覧できるかを規定するデータアクセス制限が適用され、組織のメンバー全員が協調して作業できるようになります。普遍的に適用されるセキュリティとガバナンスのコントロールにより、コンプライアンス命令が簡素化され、サイバーセキュリティのリスクが軽減されます。



ヒント

選択したデータベースオブジェクトへのライブの読み取り専用アクセスを許可することで、データをコピーしたり移動したりすることなく、Snowflakeのリージョン内で1対1、1対多、多対多のデータシェアリング関係を構築することができます。さらに、対象のアカウントで透過的にデータの同期を維持するSnowflakeグローバルレプリケーションテクノロジーを活用することで、複数のリージョンやクラウドにまたがってデータとオブジェクトを共有できます。

データクラウドを利用すると、広範な、またはパブリックなエコシステム内でのデータシェアリングも簡単に行うことができます。Snowflakeでは、他のデータマーケットプレイスが提供している「One size fits all（1つだけで全部が事足りる）」アプローチとは異なり、データプロバイダーがクライアントごとに独自のデータセットを構築することができます。Snowflakeのプラットフォームは、従来のファイルサブスクリプションサービスでは不可能なレベルのガバナンスを提供します。セキュアな行レベルのアクセスを使用して、共有するデータを指定することができます。データプロバイダーは、通路の向こう側にいるデータコンシューマーにリーチするのと同じくらい簡単に、クラウドやリージョンの境界を越えてデータコンシューマーにリーチできます。

- » Snowflakeアカウントを作成する
- » データクラウドにデータを移行する
- » セキュリティとガバナンスに注意を払う
- » リージョンを拡大する
- » データを共有して収益化する

## 第7章

# データクラウドの利用を開始するための6つのステップ

**デ**ータクラウドは、組織がデータへのアクセス、集約、統合、分析、共有を簡単かつ安全に行うことができる、データで接続された世界というSnowflakeのビジョンを実現します。Snowflakeをご利用中のお客様は、すでにこのグローバルネットワークの一員です。新規のお客様は、次のステップに沿ってご利用を開始してください。

- 1. アカウントを作成する。**まず<https://www.snowflake.com/?lang=ja>にアクセスして、選択したリージョン（通常はお望みのリージョンやクラウドプロバイダー）で最初のSnowflakeアカウントを作成します。組織内の他のメンバーが追加のアカウントを作成する場合、それらのアカウントはSnowflakeによって自動的かつ安全に接続されるため、アカウント間でデータを複製したりリージョンをまたがって簡単にコラボレーションできます。Snowflakeアカウントを新規に作成するコマンドは初期アカウントから直接実行され、管理者の役割を持つすべてのユーザーが利用できます。
- 2. データを読み込み、統合する。**データクラウドに読み込むデータソースを特定します。既存のデータウェアハウスやデータストアからデータをステージングしますか？データは連続的に読み込みますか？それともバッチモードでコピーしますか？最初のデータ読み込みとその後の段階的な更新方法を検討してください。たとえば、過去のデータセットを1回だけバルク転送するよう設定できます。新

しいトランザクションが発生したときにそのデータセットを更新する場合は、データの継続的なストリームを管理するためのデータパイプラインを構築してください。既存のアプリケーションやデータベースからデータを読み込む場合は、そのようなデータサイロからデータクラウドにデータをプッシュするソリューションを提供している商用データプロバイダーをSnowflakeデータマーケットプレイスでお探してください。一般的なデータタイプ用のコネクタとアダプターが用意されています。

3. **ガバナンスとセキュリティを適用する。** データドリブなビジネスを構築するには、データ品質を維持し、規制ガイドラインに従い、データを安全に保つための優れた管理体制が必要です。最初のデータ読み込みとその後の継続的な新規データ取得の両方において、データのガバナンスおよびセキュリティの責任者を任命してください（第2章を参照）。
4. **他のリージョンに拡張する。** 最初のアカウントから新たに追加されたSnowflakeのリージョンに簡単にアクセスすることで、データクラウドでのグローバルなプレゼンスを高めることができます。新たにリージョンを追加するたびにSnowflakeアカウントを新規作成する必要はありません。複数のリージョンからSnowflakeにアクセスできるようにすると、GDPRのデータレジデンシー要件に準拠したり、マルチクラウド戦略を実施したりする際に役立つ場合があります。
5. **データを共有する。** 自社のエコシステム内でデータを共有する予定がありますか？ そうであるなら、現在どんな方法でデータを共有していますか？ データは組織内でのみ共有する予定ですか？ それとも顧客やパートナーとも共有する予定ですか？ FTPや電子メールなどの古いデータ共有方法を特定し、それらをSnowflakeのセキュアデータシェアリングテクノロジーに置き換える方法を検討してください。
6. **Snowflakeデータマーケットプレイスに参加する。** Snowflakeデータマーケットプレイスで提供されているサードパーティのデータやデータサービスを利用したり、自社のデータやデータサービスを他のSnowflakeの利用者に提供して収益化したりする予定はありますか？ アカウントを作成してデータを読み込んだら、次の手順に従ってSnowflakeデータマーケットプレイスに参加してください。
  - コンシューマーのアクセスを許可するオブジェクトのコレクションであるシェアを特定します。
  - 1つ以上のマーケットプレイスのリストと、プロバイダーのプロファイルを作成します。
  - 125以上のサードパーティのデータプロバイダーやデータサービスプロバイダーから提供されている、すぐにクエリ可能なデータセットに安全にアクセスして、自社のデータを拡充します。

# データを活用して顧客や自社のビジネスに 最高のサービスを提供

IDCでは、2020年から2023年の3年間に作成されるデータ量が、過去30年間に作成されたデータ量を上回ると予測しています。一方、Accenture社が事業経営・IT担当専門職を対象に行った調査では、クラウドの活用によって期待した成果を完全に達成できたと答えた回答者は37%に過ぎませんでした。Snowflakeのデータクラウドは、何千もの企業がサイロ化したデータを統合し、管理されたデータを簡単に見つけて共有し、多様な分析ワークロードを実行するグローバルネットワークです。本書では、ビジネスを新たなフロンティアに導く方法をご紹介します。

## 本書の内容...

- データが生み出す課題と機会
- データクラウドでできること
- ローカルとグローバルでのデータシェアリング
- Snowflakeデータマーケットプレイス
- 産業界向けデータクラウド
- データクラウドを動かすエンジン
- 多様な分析ワークロードの実行
- 実世界におけるデータクラウドのケーススタディ



**David Baum** (david@dbaumcomm.com) は、科学とテクノロジーを専門とするフリーランスのビジネスライターです。

表紙イメージ：Snowflake提供

ビデオ、ステップバイステップの写真、  
ハウツー記事、そしてご購入は、  
**Dummies.com**™で！

ISBN: 978-1-119-86754-8  
再販禁止

for  
**dummies**®  
A Wiley Brand



# **WILEY END USER LICENSE AGREEMENT**

Go to [www.wiley.com/go/eula](http://www.wiley.com/go/eula) to access Wiley's ebook EULA.