

学びを簡単に

Snowflake 特集号第 2 版

データシェアリング

for
dummies[®]
A Wiley Brand



セキュアな
エンタープライズ
データシェアリングとは

共有データの移動に伴う
煩わしさを回避する方法

共有データを収益化
する方法

提供:



Lawrence C. Miller

David Baum

Snowflake 社について

Snowflake 社は、最新のデータウェアハウジングをすべてのデータユーザーにとって効果的かつ安価で、アクセスしやすいものにするという明確なビジョンを持って創業しました。Snowflake 社は、複数のクラウドにわたって即座の弾力性、セキュアなデータシェアリング、秒単位の価格設定を提供することでデータ主導の企業を実現します。従来のオンプレミスおよびクラウドソリューションはこの点に苦戦していたため、Snowflake 社は、データウェアハウジングのパワー、ビッグデータプラットフォームの柔軟性、クラウドの弾力性を併せ持つ、クラウド向けに構築された新しいアーキテクチャを備えた新製品を開発し、従来のソリューションの数分の1の価格で提供しています。Snowflake : データに限界はありません。

詳細については、Snowflake 社のサイト ([snowflake.com](https://www.snowflake.com)) をご覧ください。



データ シェアリング

Snowflake 特集号第 2 版

**Lawrence C. Miller、
David Baum 共著**

for
dummies[®]
A Wiley Brand

データシェアリング For Dummies®、Snowflake 特集号第 2 版

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、録音、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一般的な情報、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) までお問い合わせいただくか、www.wiley.com/go/custompub をご覧ください。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 までお問い合わせください。

ISBN 978-1-119-84433-4 (ペーパーバック) ; ISBN 978-1-119-84435-8 (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。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、メール (info@dummies.biz) にてお問い合わせいただくか、www.wiley.com/go/custompub をご覧ください。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 までお問い合わせください。

本書の出版にあたりご協力いただきました皆様により心より御礼申し上げます。

開発編集者: Nicole Sholly

事業開発担当: Karen Hattan

プロジェクトエディター: Martin V. Minner

プロダクションエディター: Siddique Shaik

エグゼクティブエディター: Steve Hayes

Snowflake 寄稿者: Vincent Morello、

Daniel Kuperman、Leslie Steere、

編集責任者: Rev Mengle

Michael Nixon、Matthew Glickman

目次

はじめに	1
本書について	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 データシェアリングの基本を押さえる	3
データシェアリングとは	4
データシェアリングの例	7
組織のデータ共有方法	7
データシェアリングの可能性を探る	8
第 2 章 従来型データシェアリングの課題を理解する	9
多面的な問題に対処する	9
従来型データシェアリングに対応する：価値実現までの遅れ	12
コンピューティングの複雑性の課題	15
従来方式のデータシェアリング：ビジネス上の問題点	15
組織内でのデータシェアリング	16
企業間（B2B）でのデータシェアリング	16
データの収益化	16
第 3 章 データシェアリングのビジネス価値を認識する	17
初期のデータ共有を振り返る	17
組織にとってのデータシェアリングの価値を評価する	18
基幹業務間でのデータシェアリング	18
組織間での共有：外向き	19
組織間での共有：内向き	19
データを収益化する	21
第 4 章 データシェアリングをライブで実現する	23
最新型データシェアリングを探る	23
最新のデータシェアリングで成功を収める	24
データシェアリングを容易にする	25
最新型クラウドデータシェアリングを活用する	29
最新型クラウドデータプラットフォームの仕組みを知る	29
セキュアビューを使用して共有データへのアクセスを制御する	30

第 5 章	最新型データシェアリングの影響を評価する	33
	データコラボレーションへの最新のアプローチ	33
	コンピューティングの民主化	34
	データ主導のビジネスの意思決定.....	35
	データの商業化.....	36
第 6 章	エクステンジを活用してデータを共有し、 インサイトを収益化する	39
	データマーケットプレイスとデータエクステンジの歴史を 検討する	39
	リアルタイムのデータシェアリングで最新化を推進する	41
	データエクステンジに適したアーキテクチャを構築する	42
第 7 章	データを収益化する	45
	業界の機会を検討する	46
	データ収益化のトレンドに正面から取り組む.....	47
	優れたデータガバナンスの遵守	48
	販売チャネルを評価する.....	48
	データシェアリングサービスの価格を設定する	49
第 8 章	データを統制する	51
	同じデータを 1 つまたは複数のコンシューマーと共有する.....	52
	データのさまざまなサブセットを複数のコンシューマーと 共有する.....	53
	セキュアビューを使用して定義済みのデータスライス を提供する.....	54
	データプライバシーを維持しながら質問に答える	55
	セキュアな結合を使用して企業間のデータを結び付ける.....	56
第 9 章	最新のデータシェアリングでビジネスを前進 させるための 6 つのステップ	57

はじめに

企業内の組織は、互いに共有するデータを分析することで重要なインサイトを獲得します。たとえば、財務チームであれば、将来の財務実績を予測するために販売データが必要です。製品管理チームであれば、将来の製品およびサービスを判断するためにマーケティングデータが必要です。経営陣であれば、ビジネス上の意思決定をデータに基づいてタイムリーに下すために、企業のさまざまな部門からのデータが入力された最新のダッシュボードが必要です。

組織の外部では、小売業者が在庫とサプライチェーンを管理するためにベンダーと販売データを共有しています。SaaS（サービスとしてのソフトウェア）プロバイダーは、収集したデータを顧客と共有し、自社のビジネスに関する詳細なインサイトを顧客に提供しています。医療機関は、補助製品を提供しているベンダーや、データを分析して患者サービスの改善に役立っている他のビジネスパートナーと患者データを安全な方法で共有しています。他にも数えきれないほどの例があります。Forrester Research 社によると、世界のデータおよび分析意思決定者の半数以上が、所属する企業が外部データの調達能力を拡大していると報告しています。

データは、単に収集して分析するだけのものではなくなりました。業務を効率化したり、パーソナライズされたカスタマーエクスペリエンスを迅速に提供したり、新しい市場機会を開拓したりするために組織の内外で簡単かつ安全に利用できる資産なのです。データプロバイダーであれば、自社のデータを安全に収益化し、無数のデータコンシューマーとセルフサービスの関係を築くこともできます。実際、Forrester 社によると、世界のデータおよび分析意思決定者の 47% が、所属する組織が現在データの商業化を行っていると報告しています。

本書について

Snowflake 特集号第 2 版『データシェアリング For Dummies』へようこそ。本書では、最新型データシェアリングを活用し、管理されたセキュアな方法で、数分以内でライブデータを共有・受信する方法について説明します。従来のデータ共有手法の悩みの種だったリスク、コスト、心配事、遅延はほとんどなくなります。最新型データシェアリングでは、1対1、1対多、多対多のデータ共有関係を新しい独創的な方法で簡単かつ迅速に構築し、インサイトを獲得するまでの時間を以前ならまったく不可能だったレベルに短縮できます。

本書で使用するアイコン

本書では、重要な情報への注意を呼び掛ける特殊なアイコンを適宜使用します。各アイコンがある箇所では取り上げる内容は以下のとおりです。



ケーススタディ

ケーススタディでは、最新のデータ共有手法を駆使して成功を収めている組織のベストプラクティスを紹介しています。



ポイント

このアイコンは、頭に焼き付けておく必要がある重要な情報を示しています。



技術情報

このアイコンがある箇所では、高度な専門用語を詳しく解説しています。



ヒント

このアイコンは、有益な情報と役に立つアドバイスを示しています。



注意

このアラートがある箇所では、コストがかかるミス、またはストレスや負担となるミスを回避するための実用的なアドバイスを提供しています。

本書を読み終えた後で

本書を読み終えた後、「もっと詳しく知るにはどうしたらよいか」と思われた方は、<https://www.snowflake.com/?lang=ja> にアクセスしてください。Snowflake へのお問い合わせ、Snowflake のテクノロジーを無料で試すこともできます。Snowflake が提供している製品やサービス、最新のデータシェアリングに関する情報、ウェビナーの視聴、今後のイベントに関する情報や、ドキュメントやその他のサポートにアクセスすることができます。

- » データシェアリングを定義する
- » データシェアリングの重要性を認識する
- » データシェアリングの事例を探る
- » 組織がデータを共有する方法を理解する
- » データシェアリングの機会を活用する

第1章

データシェアリングの 基本を押さえる

毎日、あらゆる分野の組織がビジネス成果の追跡、意思決定、顧客とのエンゲージメント、製品の定義と作成、トレンドの予測などにデータを活用しています。データは、互いに内部および外部の複数の組織が事業計画、相互イニシアティブ、合併事業などでコラボレーションする際に利用および消費されるリソースでもあります。

企業がデータを活用して連携・コラボレーションする方法は無限にあります。ただし、比喩的に言うと、データは魔法のように玄関先に現れるものではありません。その発生場所で生成され、その後、組織全体に配布され、インサイトを得るために分析されるものなのです。

本章では、データシェアリングについて説明します。データシェアリングとは何でしょうか。データシェアリングが重要なのはなぜでしょうか。組織はなぜ、どのようにしてデータを共有するのでしょうか。そしてデータシェアリングによってどのようなビジネスチャンスが生まれ出されるのでしょうか。

データシェアリングとは

データは、企業がビジネスの運営に利用する数多くのソフトウェアアプリケーション、Web サイトにアクセスする訪問者の持続的な活動、家庭の冷蔵庫に取り付けられたIoT（モノのインターネット）デバイス、旅客機のジェットエンジンのような高度なものに搭載されたセンサーなどから生み出されます。現代社会では、データを作成するシナリオは無限にあると言っても過言ではありません。市場調査会社の IDC は、全世界で創出されるデジタルデータの総量が 2025 年には 180 ゼタバイトにまで増加すると予測しています（1 ゼタバイトは約 1 兆ギガバイトに相当）。残念ながら、従来のデータシェアリング方法では、非常に問題の多いデータ移動が必要になります。今後、膨大な量のデータを有意義なやり方で共有することは、不可能ではないにせよ、現実的とは言えないでしょう。

さらに、多くの企業は、組織外のデータへアクセスできれば事業運営を強化できることに気が付き始めました。また、自社で生成していないデータへのアクセスが容易でないことも認識しています。したがって、データシェアリングとは、同じ企業内のビジネスユニット間または互いに外部の存在である組織間でデータへのアクセスを提供する行為を指します。自らが持つデータを共有する組織はデータプロバイダーと呼ばれます。共有されたデータを利用する組織はデータコンシューマーと呼ばれます。どのような組織も、データプロバイダー、データコンシューマー、またはその両方になる可能性があります。

図 1-1 は、組織の従来へのデータシェアリング方法を示しています。この方法では、組織は共有データのコピーを作成し、それをデータコンシューマーに送信します。その後、データコンシューマーはデータをダウンロードして分析したり、社内の既存データと組み合わせたりして、顧客の素性、事業運営の効率性、新たに参入すべき産業などに関する詳細なインサイトを獲得します。

しかし、このプロセスは面倒で、時間もコストもかかり、限られた量の共有データしか移動できません。図 1-2 は、データ移動のない最新のデータシェアリング方法を示しています。この方法では、データプロバイダーは、最新のクラウドデータシェアリングを介してデータのライブの読み取り専用コピーをデータコンシューマーに提供します。本質的にデータ移動は不要です。



ポイント

最新のデータシェアリングは、最新のデータプラットフォームに搭載されている機能です。

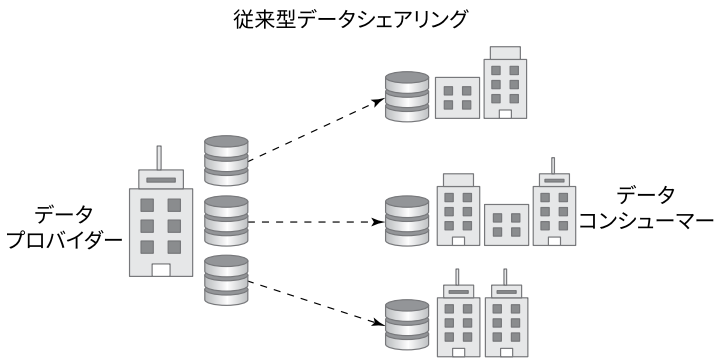


図 1-1: 従来型データシェアリングでは、データを複製し、データプロバイダーからデータコンシューマーに移動する必要がある。

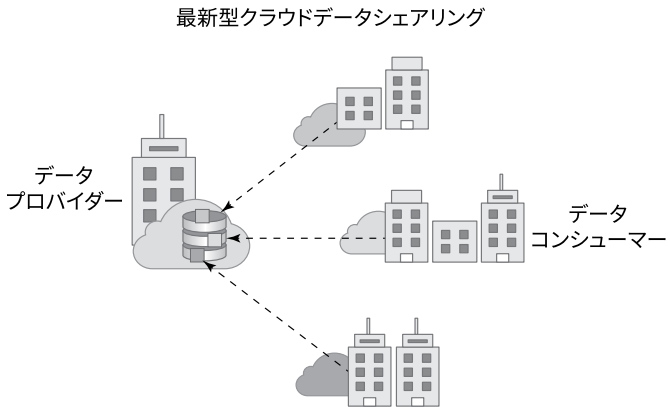


図 1-2: 最新のクラウドデータシェアリングでは、データを移動することなく、高速、ライブ、かつ安全に、管理された環境下でデータを共有できる。

求職者向けポータル、効率性を 300% 向上



ケーススタディ

Snagajob 社のモバイル調達・採用ツールは、7,500 万人の登録済み時間給労働者と同社の法人サブスクライバー（30 万の雇用場所に対応）を結び付けています。

Snagajob 社は、同社のデータプラットフォームのデータを外部のマーケティング分析サービス企業と共有する必要がありました。この企業は、Snagajob 社の企業顧客に接触し、Snagajob 社に代わって対象を絞り込んだ再エンゲージメントキャンペーンを実施するためにその共有データを利用しました。Snagajob 社は、この企業との関係により、このマーケティング機能に関連する社内人件費を回避できました。

Snagajob 社は、データを共有するために、次のような時間のかかる手順を日常的に実施する必要がありました。

- 共有するデータベース要素を特定する
- クライアントツールを使ってデータセットを抽出する
- データを E メールで送信できるように圧縮・暗号化する
- マーケティングパートナーにファイルを E メールで送信する

外部のマーケティング企業は、ファイルを受け取った後、同様に時間のかかる手順を逆の順序で実行する必要がありました。さらに、データを取り込むためのデータベーステーブルを作成し、データをそのターゲットデータベースにインポートするという手順も必要でした。

そこで Snagajob 社は、非常に拡張性の高いデータウェアハウスとしてだけでなく、データ共有のプラットフォームとしても機能できる、クラウド向けに構築された最新のデータプラットフォームを採用しました。Snagajob 社が行う必要があったのは、Snagajob 社が共有したテーブルとビューにマーケティングパートナーがライブで安全かつ同時にアクセスできるようにする「共有」を作成することだけでした。マーケティングパートナーは、提供されたデータを使って E メールキャンペーンを実施できました。データは常に正確かつ最新でした。パートナーは常に Snagajob 社の最新データのライブの読み取り専用バージョンにアクセスできたからです。このプロセスはほんの数分で完了し、データ移動は一切発生しませんでした。

パフォーマンス、信頼性、アジリティが飛躍的に向上し、Snagajob 社はデータシェアリングの実装に要する時間を数日からわずか数時間に短縮できました。Snagajob 社はコストを 300% 削減し、業務をこれまでよりもはるかに迅速に行えるようになり、業界での競争力を高めました。

データシェアリングの例

以下に、最新のクラウドデータシェアリングによって可能になる新しいビジネスチャンスいくつかをご紹介します。

- ▶▶ **データサイロ解消手段としてのデータシェアリング**：すべての社内データを統合した信頼できる唯一の情報源を構築し、それを1つの組織内に何百とあるビジネスユニットの何千ものデータコンシューマーで共有します。
- ▶▶ **業務効率化手段としてのデータシェアリング**：ビジネスパートナーとライブのデータを共有することで、コストの最適化、業務の合理化、優れた顧客サービスを実現します。
- ▶▶ **製品としてのデータシェアリング**：データコンシューマーが既存のデータを拡充できるように、データスライスへのライブの直接アクセスを有料サービスとして提供します。
- ▶▶ **製品差別化手段としてのデータシェアリング**：SaaS（サービスとしてのソフトウェア）プロバイダーは、B2B（企業間）サブスクライバーのアクティビティから生成される数ペタバイトのデータへの直接アクセスを提供できます。これにより、サブスクライバーは、これまで実行できなかった大量データに対する詳細な分析を実行できるようになります。

上記のような機会を活用するには、並外れたスピード、パワー、セキュリティ、ガバナンス、そしてシンプルさを兼ね備えたデータシェアリング機能が必要になります。これらの機能は、従来のデータシェアリング方法では利用できません。

組織のデータ共有方法

従来のアプローチでデータを共有および移動する場合、バラバラなタスクをパッチワークのようにつなぎ合わせる必要があります。以下のようなプロセスはコストがかかり、手作業によるオーバーヘッドを伴い、組織が実際に共有できるデータの量を制限します。

- ▶▶ **Eメール**：データファイルは、プロバイダーからコンシューマーにEメールで送信されます。
- ▶▶ **ファイル転送プロトコル (FTP)**：データファイルは、2台のコンピュータ間で、またはインターネット経由で共有・ダウンロードされます。

- » **ETL (抽出・変換・ロード) ソフトウェア**：ETL ソフトウェアがプロバイダーのデータベースからデータを抽出して変換し、コンシューマーのデータベースにロードします。
- » **オンラインファイル共有サービス**：FTP と似ていますが、データファイルの共有とダウンロードはインターネットファイル転送のみを介して行われます。
- » **クラウドストレージ**：プロバイダーがデータをクラウドに保存し、そのデータにアクセスするための認証情報をコンシューマーに提供します。
- » **アプリケーションプログラミングインタフェース (API)**：データ転送を開始および管理するために API が使用されます。

しかし、想像してみてください。すぐに利用可能なライブデータにオンデマンドでアクセスでき、そのデータを安全な管理された環境で即座に利用できるとしたらどうなるでしょうか。

データシェアリングの可能性を探る

最新のデータシェアリングの可能性は事実上無限と言っても過言ではありません。以下に、魅力的な機会をいくつかご紹介します。

- » **カスタマーエクスペリエンスを向上させる**：競争の激しいデジタルマーケットでパーソナライズされたマーケティングキャンペーンを活用して対象を絞り込んだビジネスまたは小売サービスを提供するには、顧客、競合他社、そして業界のトレンドを十分に理解しておかなければなりません。このような理解を得るための一番の近道は、まだ知らないことを明らかにするためにまだ手元にないデータを手に入れることです。
- » **業務を効率化する**：組織を構成する多数のビジネスユニットやビジネスパートナーとデータを容易に共有できると、信頼できる唯一の情報源が創出されます。これにより、ほんの些細なデータ不整合を修正することで何十億ドルものコストを削減できます。
- » **データから新しいビジネス資産を創出する**：データプロバイダーが保持しているデータの中には、外部の競合しない何千ものデータコンシューマーにとっても同様に有益なものがあります。最新のデータシェアリングを提供する最新のクラウドデータプラットフォームでは、任意のデータ要素を簡単に共有できるため、このような大規模な共有を手間のかからないセルフサービス方式のビジネスモデルで実現できます。

本章の内容

- » 現在のデータシェアリング方法を確認する
- » 遅延のコストを認識する
- » コンピューティングの複雑さを確認する
- » ビジネス上の問題に対処する

第2章 従来型データシェアリングの課題を理解する

本章では、従来型データシェアリングの方法やテクノロジーのさまざまな制限について説明します。

多面的な問題に対処する

組織の内外に何十、何百、何千とおり、それぞれに固有のデータシェアリング要件があるデータコンシューマーとデータを共有することを想定してみてください。この課題に容易に対応できる方法はあるのでしょうか。ストレージクラスターを絶えず増設したり、複雑なソフトウェアを管理したり、長時間の遅延やパフォーマンスの低下に悩まされたりせず、また古いデータを共有することによる不整合を生じさせたりせずに、成長を支えるにはどうすればよいのでしょうか。端的に言えば、現在の従来型データウェアハウスプラットフォームは、リアルタイムでデータを共有するという不断の必要性に対応するようには作られていません。

従来型のデータシェアリングが抱える課題がいかに大きいかを理解するため、企業がサードパーティのサービスプロバイダーや、ビジネスパー

トナーなどの別の外部組織とデータを共有する場合などに遭遇するであろう一般的アプローチの利点と欠点（表 2-1）を考えてみましょう。

表 2-1 従来型データシェアリングの利点と欠点

データシェアリング手法	利点	欠点
Eメール	<ul style="list-style-type: none"> ● 広く普及し、誰もが利用している ● インフラが整備されている ● Eメールの作成やファイルの添付が容易 	<ul style="list-style-type: none"> ● リレーショナルデータベースの大規模なデータセットには不向き（拡張性に乏しい） ● 添付ファイルにサイズ制限がある（25 MB 未満に限る）ため、大規模なデータセットは分割して圧縮する必要がある ● ネットワーク帯域制限によりデータ転送速度が遅くなる ● 安全性が低く、独自の暗号化を要する ● 受信側でミラーリング作業（データの受信、復号化、再構築など）が必要
ファイル転送プロトコル (FTP) — 図 2-1 参照	<ul style="list-style-type: none"> ● 広く知られており、長く利用されているプロトコル ● さまざまな FTP クライアントソフトウェアやサービスが利用可能 	<ul style="list-style-type: none"> ● スキーマの変更に多大なリードタイムを要する ● FTP 用のクライアントソフト、サーバー、サービスを手入する必要がある ● FTP アカウント管理者を設定する必要があり、オーバーヘッドが生じる ● 大規模データセットの場合、データ転送を高速化するために分割してサイズを縮小する必要がある ● ネイティブにセキュアでないため、カスタムの暗号化スクリプトやセキュアなサービスを要する ● 受信側でミラーリング作業（データの受信、復号化、再構築）が必要 ● 共有データセットを更新するたびに作業を繰り返す必要がある

データシェアリング手法	利点	欠点
ETL（抽出・変換・ロード）ソフトウェア	<ul style="list-style-type: none"> ● 定評のある ETL ソフトウェアが豊富に利用できる ● データベースやデータソースからデータを抽出し、変換してターゲットデータベースにロードすることに特化した設計 ● 大量かつ複雑なデータの移動や変換に適している 	<ul style="list-style-type: none"> ● データが変更されると遅延が発生する ● ソフトウェアが高価で、最大数百万ドルのコストがかかる ● 複雑で、統合や導入に専門的スキルを要する ● 実装に数か月かかる場合がある ● 変更管理やスキーマ進化が困難な場合がある
オンラインファイル共有サービス	<ul style="list-style-type: none"> ● サービスの可用性が高い ● 一般的に使いやすい 	<ul style="list-style-type: none"> ● フラットファイルの共有には適しているが、リレーショナルデータベースのオブジェクトには適さない ● データをすぐに利用できない（分析できない） ● データのオリジナルコピーが変更されると、データの不整合に伴うリスクが発生する
クラウドストレージ	<ul style="list-style-type: none"> ● 大手クラウドストレージプロバイダーが提供するサービスが多数ある 	<ul style="list-style-type: none"> ● クラウドストレージから直接クエリを実行した場合、最適なパフォーマンスが得られない ● 変更管理やスキーマ進化が難しく、別途メタデータ管理のプロセスを要する ● データ変更時にリスクが発生する ● 完全な SQL データ操作言語（DML）セマンティクス（UPDATE、INSERT など）がサポートされていない場合がある

（続く）

表 2-1 (続く)

データシェアリング手法	利点	欠点
アプリケーションプログラミングインターフェイス (API)	<ul style="list-style-type: none"> ● 利用可能な API が多数ある ● 多様なユースケースがある ● プログラムによる実装で、手動での作業負担を軽減できる 	<ul style="list-style-type: none"> ● データ移動が必要なため、転送失敗のリスクがある ● API では少量のデータを処理するため、大量のデータ処理を行う場合にボトルネックとなる ● パフォーマンスが利用可能な帯域幅に直接影響されるため、高帯域幅を利用するには高いコストを要する

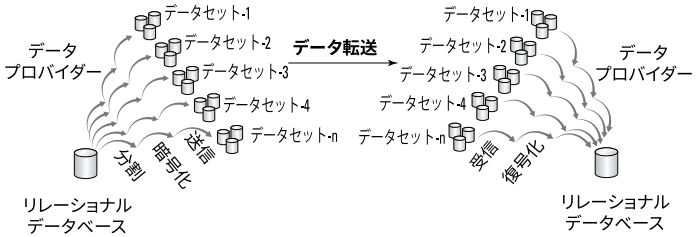


図 2-1: 典型的な旧式の FTP ベースのデータシェアリングワークフローにおける複数のステップ。

従来型データシェアリングに対応する：価値実現までの遅れ

従来型のデータシェアリングは、さらなる遅延を引き起こし、IT チームからの追加のサポートを必要とするその他の課題を生み出す可能性があります。このような課題として以下が挙げられます。

- » **データサイズの増大への対応**：共有データセットは、当初の想定よりはるかに大きくなるがよくあります。これにより、データ抽出プロセスで問題が生じます。このような問題が発生すると、分割・抽出プロセスを自動化するスクリプト言語が必要になる可能性があります。この場合、追加の IT サポートが必要になることがあります。また、データコンシューマー側でも逆のプロセスを実行する必要があります。
- » **機密データの復号化**：データセットに機密情報が含まれている場合、出力ファイルの暗号化、マスク処理、または墨消しが必要になる可能性があります。この場合、追加の IT サポートが

必要になることがあります。データセットを暗号化する場合、別のプロセスを使用して暗号化キーを当事者間で安全に共有する必要があります。また、データコンシューマー側で共有データを復号化する必要があります。

- ▶ **ファイル形式やスキーマの変更**：追加のデータベース属性を共有する必要がある場合は、ファイル形式を複数回変更することが必要になる場合があります。データプロバイダー側でテーブル属性が変更されると、データコンシューマー側でもデータを変更する必要があります。

これらのステップが積み重なり、データプロバイダーとデータコンシューマーの両方でプロセスが時間と手間のかかるものになります。また、これらの作業はデータを分析してインサイトを獲得する前に行っておく必要があるため、価値実現までの時間が長くなります。

一般に、これらの遅延と困難はデータ転送作業だけでなく、以下のような作業でも発生します。

- ▶ **リアルタイムでのデータ共有**：データセットを一晩に一度だけ送信するのではなく、よりリアルタイムに近い方法で共有する場合には、追加の IT サポートが必要になります。
- ▶ **データのクリーニング**：インポートプロセスに問題があり、データが予想よりもクリーンでないことがあります。たとえば、抽出データに本来なら無視されるべき特殊文字が含まれている場合があります。これはデータプロバイダーがより洗練されたデータ抽出プロセスを構築する必要があることを意味しており、その結果、必要な IT サポート、コスト、遅延が増加します。

数十億台のデバイスからの データアクセスを高速化



ケーススタディ

ボストンに拠点を置く Localytics 社は、世界中の 27 億台以上のモバイルデバイスの 37,000 以上のアプリで使用されているモバイルエンゲージメントプラットフォームを提供しています。Localytics 社は、そのモバイルユーザーに関するインサイトと、それらのユーザーと関わり合うためのツールを世界のトップブランド数百社に提供しています。

Localytics 社は、最新のデータシェアリングを活用し、顧客がデータをエクスポートせずに Localytics 社のデータにアクセスできるようにすることで、マーケティング担当者が直面する最も大きな課題の 1 つを解決しています。以前はユーザーは、顧客関係管理 (CRM) システム、ビジネスインテリジェンス (BI) ツール、モバイルアナリティクスなどのソー

(続く)

(続く)

スから得られたさまざまな顧客データのソースを結び付ける必要がありました。また、データをエクスポートして他のプラットフォームにコピーし、そこでデータを操作したり、さらなる分析を行ったりすることがよくありました。このプロセスにより、データの複数のコピーがさまざまな場所に生成されました。その結果、コストと複雑さが増すとともに、整合性のある結果を得ることができませんでした。

ETL の排除

Localytics 社は、手間のかかる ETL 作業の負担をなくし、最新のデータシェアリングを介してデータに直接アクセスするようにしました。これにより、これまでよりもはるかに効率的に、かつ信頼性の高い方法で顧客データを管理・把握できるようになりました。具体的に言うと、同社はセキュアな権限ベースのアクセスを採用し、顧客が同社のセッション、イベント、プロフィールデータを操作したり、それらのデータに対して独自のクエリやカスタムレポートを実行したりできるようにしています。顧客は、一般的な BI ツールを使用してデータを分析することもできます。

データレイテンシーを 3 時間から 3 分に短縮

Localytics 社は、最新のデータシェアリング機能を備えた最新型クラウドデータプラットフォームにすべてのデータを保存しています。顧客とライブデータを即座に共有することで、以前の 3 時間というデータレイテンシーを解消しました。最新のデータシェアリングにより、リアルタイムデータを約 3 分で照会できるようになりました。顧客側では、データを活用するための追加作業は必要ありません。データは、ライブかつセキュアで、管理された、権限ベースのシェアリングによってフィルタリングされます。

顧客は、ETL プロセスを排除できる可能性があることにすぐに気づき、過労のデータチームからその負担を取り除くことと組織の時間と労力を節約することに成功しました。

抽出側またはインポート側のいずれかでファイル転送プロセス中に障害が発生するのを防ぐには、データプロバイダーとデータコンシューマーの両方が、転送を監視し、障害が発生した場合にプロセスを自動で再開する特別なソフトウェアコードまたはスクリプトを組み込む必要があります。これは、インサイトを獲得してデータから価値を引き出すのに要する労力と時間が増えることを意味します。

コンピューティングの複雑性の課題

また、従来型のデータシェアリングオプションでは、ほんの小さなデータスライスを共有する場合でさえも複雑なコンピューティングプラットフォームを拡張する必要があります。複雑さによって負担が増し、組織の内外でインフラストラクチャコストを含む追加のリソースが必要になります。

目標は、ビジネスパートナーを含む内外の組織と無限の量のデータを共有し、コラボレーションや事業計画に活用するための容易なデータシェアリングを実現することです。データの収益化に重点を置いたビジネスモデルを採用している場合は、必要に応じて個別化されたセルフサービス方式のアクセスとセキュリティを活用してできる限り多くのデータコンシューマーにデータを配布するための、これと同程度に容易なデータシェアリングが必要になるでしょう。



注意

クラウドストレージが解決策になるとお考えの場合は、考え直してください。基本的なクラウドストレージサービスを利用したデータシェアリングは非効率的です。そのようなデータシェアリングでは、組織とそのデータコンシューマーは高性能にデータをクエリしたり、データの整合性を確保したりすることができません。Hadoop コンピューティングプラットフォームは、その複雑さゆえに解決策とはなり得ません。

従来方式のデータシェアリング： ビジネス上の問題点

面倒で複雑なデータシェアリング手法は、高額で柔軟性に欠けるコンピューティングプラットフォームと相まって、データを活用したコラボレーションを必要とする組織にとって悩みの種となります。また、従来型のデータプラットフォームからデータを抽出し、それを他の組織へ転送するために必要なオーバーヘッドにより、共有データからの価値を獲得するのに遅延が生じます。さらに、共有データは常に静的バージョンなので、すぐに陳腐化するため、データを変更するたびにデータの抽出・転送プロセスを繰り返す必要があります。

データから素早くインサイトを抽出できないという点は、データから最大限の商業的価値を引き出すことを阻む要因となっています。データコンシューマーはインサイトを得るまでに時間がかかり、それはデータプロバイダーへの不満へとつながります。さらに、一般的なデータシェアリング手法では、変更内容をすぐ反映させることができないため、データコンシューマーは不完全なデータを対象に分析を行ってしまうというリスクを負います。これは、正確性に劣る分析やビジネス上の誤った意思決定につながる可能性があります。

組織内でのデータシェアリング

組織内で想定されるデータシェアリングのシナリオとして、以下のよう
なものが挙げられます。

- ▶▶ 営業部門が財務部門とデータを共有し、売上と収益を追跡して組織の業績を予測する。
- ▶▶ マーケティングチームが顧客データを監視・分析して顧客行動を予測し、需要創出プログラムを調整する。
- ▶▶ 組織のさまざまな子会社が相互にデータを共有して市場開拓計画を改善し、それぞれの事業分野に関する理解を深める。

組織内の各部門がデータを効率良く共有できない場合、データサイロが発生し、ビジネスコラボレーションに支障をきたします。この場合、各チームが独自のデータウェアハウスやデータマート（会社のデータウェアハウスに保存されているデータの一部のコピー）を保持することになります。その結果、データサイロとデータマートが無秩序に増加し、ITチームとデータプラットフォームチームに余分な負担がかかります。

企業間（B2B）でのデータシェアリング

B2Bでのデータシェアリングの例として、以下のようないものが挙げられます。

- ▶▶ ホテルの予約ウェブサイトですべての予約のパターンや傾向をホテル側と共有し、プロモーション・価格設定プログラムを開発する。
- ▶▶ 食料品チェーンが店舗の売上データをサプライヤーに提供し、需要に応じた適切な在庫を確保できるようにする。
- ▶▶ 小売業者が店頭販売データをファッションマーチャンダイジング担当者として共有し、最も人気のあるトレンドアイテムを常に確保できるようにする。

外部組織に共有データを送る場合であれ、外部組織から共有データを受け取る場合であれ、データを活用したコラボレーションができれば、効率性が低下し、運用コストの上昇と生産性の低下を招くリスクを負います。

データの収益化

共有データの収益化の例として、携帯電話の位置情報や利用状況データを収集して広告代理店やマーケティング部門と共有し、それらの組織が特定の消費者に向けて高度にターゲットを絞ったキャンペーンを実施できるようにするデータサービス企業が挙げられます。

データの収益化については、第3章で詳しく説明します。

- » ビジネスにおけるデータシェアリングの歴史をたどる
- » データシェアリングのシナリオとビジネスにもたらす価値を詳しく検討する

第3章 データシェアリングの ビジネス価値を認識する

本章では、ビジネスにおけるデータシェアリング手法の進化、あらゆるビジネスでデータシェアリングが重要な理由、企業が社内外でデータを共有する方法、クラウドとSaaS（サービスとしてのソフトウェア）がデータシェアリングモデルにもたらした変化について説明します。

初期のデータ共有を振り返る

今日のデータシェアリングのビジネス価値を理解するには、その歴史を振り返る必要があります。少し前まで、企業が自社のデータセンター内で複数のビジネスアプリケーションをホストしてサポートするのはごく当たり前のことと考えられていました。アプリケーションには、財務用、マーケティング用、営業用、人事用、業務用などさまざまなものが存在します。10年前であれば、大企業は何百というビジネスアプリケーションを自社のデータセンターでホストし、実行していたでしょう。

これらの各アプリケーションにはデータベースも関連付けられました。それらのデータベースは、分析用に最適化されておらず、アプリケーショ

ン間でデータを共有していませんでした。このデータを分析するには、データベースを担当する各事業部門が、データウェアハウスの小型スタンドアロン版である独自のデータマートにデータを抽出、変換、ロード（ETL）する必要がありました。次に、組織全体でビジネスインテリジェンスを開発したり、全社的なデータを対象に分析を実行したりするために、個々のデータマートから中央のデータウェアハウスに ETL プロセスを介してデータを送信する必要がありました。その後、このデータを分析用に準備しました。このプロセス全体は時間も手間もかかるものでした。データ形式はアプリケーションごとに異なっていたため、新しいデータウェアハウスに適合するようにさらなるモデリングと変換を行う必要がありました。しかし、少なくともデータにアクセスすることはできませんでした。データが自社のデータセンターにあったからです。

結論として言えることは、どんな企業も社内である程度のデータシェアリングを行わない限り生き残れないということです。

組織にとってのデータシェアリングの価値を評価する

組織内全体および複数の組織間でのデータシェアリングは、以下の4つの基本ワークフローで構成されます。

- ▶▶ **基幹業務 (LOB) 全体:** 同じ組織内の複数の事業部門間でのデータシェアリング
- ▶▶ **組織間:** 自社にとって利益となる、別の組織への外向きのデータシェアリング
- ▶▶ **組織間:** 自社にとって利益となる、別の組織からの内向きのデータシェアリング
- ▶▶ **データ収益化:** データコンシューマーがその既存データを拡充できるようにライブデータをサービスとして共有すること

基幹業務間でのデータシェアリング

同じ組織内の各事業部門は、Eメール、スプレッドシート、共有ネットワークドライブ、アプリケーションプログラミングインターフェイスなどのさまざまな方法を使用してコミュニケーションやデータの共有を行っています。組織全体でデータを共有すると、ビジネスインテリジェンスのレベルの向上が実現・促進されるとともに、タイムリーで情報に基づいた意思決定が促進されます。

しかし、組織内では往々にして、データがサイロに閉じ込められています。合併や買収、ファイアウォールによる制限などのビジネス上または技術上の障壁により、組織の事業部門間でのデータの共有が困難になることがよくあります。このようにインフラストラクチャが物理的または論理的に分離されている場合、複数の事業部門が組織内の利用可能なすべてのデータにアクセスし、包括的でデータ主導型のインサイトを実現することができなくなる可能性があります。このようなデータサイロは、組織が従来型のオンプレミスデータウェアハウスやクラウドに移行した従来型データウェアハウスに依存している場合に発生します。

組織間での共有：外向き

外部とのデータ共有は常に行われています。ベンダーとサプライヤーの関係、パートナー関係、開発者と生産業者の関係、それ以外にもさまざまなビジネス関係があり、どの関係においても複数の組織がビジネスを推進するためにデータを共同で活用する必要があります。図 3-1 は、主たる組織がパートナー組織に対して外向きにデータを共有する例を示しています。

たとえば、ベンダーとサプライヤーがデータを共有する関係では、サプライヤーが特定の商品の在庫を補充するタイミングを事前に知ることが

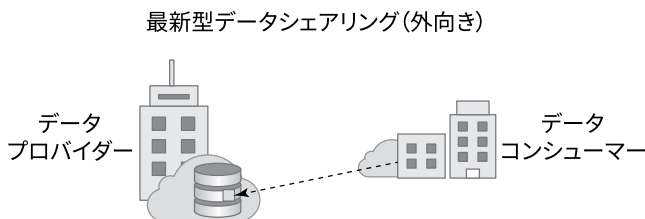


図 3-1: データプロバイダーの役割を果たす組織がデータコンシューマーとなるサプライヤーとデータを共有する。

できます。また、在庫が適切に管理されている場合は、過剰在庫を防ぎ、価格を大幅に下げる必要性を最小限に抑えることができます。

組織間での共有：内向き

組織は外部のサービス企業と提携する機会がますます増えています。このような提携先企業は、物流、配送、マーケティングサービス、販売業務などを専門としています。たとえば、大手小売業者は、そのターゲット顧客に関する膨大な量の人口統計データを収集します。その後、データプロバイダーとしてデータ分析企業とデータを共有します。

データシェアリングによって社内業務を効率化



ケーススタディ

楽天リワードは、1997年以來、人々のオンラインショッピングのやり方に変化をもたらしている国際的複合企業であり、世界最大級の品揃えを誇る商品やサービスを対象にキャッシュバックやショッピングポイントを提供しています。同社の各部門または各子会社には、共有可能なデータに関連する特定の法的要件と権限があり、社内のデータ共有インフラストラクチャが複雑化していました。

社内のデータ共有プロセスが煩雑になっていただけでなく、各事業部門が情報に基づく意思決定を下すために互いのデータセットにアクセスすることもできませんでした。

楽天は、最新のクラウド構築型データプラットフォームを採用しました。これにより、同社の事業部門と子会社は、管理された共有データを数分で簡単に取得し、それを独自のデータソースと組み合わせるより詳細なインサイトを得られるようになりました。

楽天の分析担当副社長を務める Mark Stange-Tregear 氏は、社内のデータシェアリングの変革について次のように述べています。「当社が当社の法的な傘の範囲外にあるグループと簡単にデータを共有できるようになっただけでなく、特定の専門知識やニッチ分析スキルを持つ他のグループが当社のデータセットを活用できるようになりました。その結果、当社は外部のリソースを雇う必要がなくなり、より機敏な組織になりました」。

データ分析企業は、小売業者に代わってデータ分析を行います。その後、図 3-2 に示すように、内向きのデータ共有という形で小売業者に分析結果を返します。

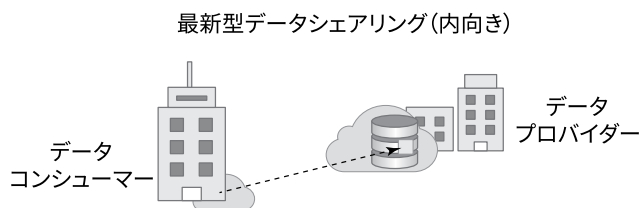


図 3-2: この組織はデータコンシューマーであり、データコンシューマーである外部のデータ分析ベンダーのデータにアクセスする。

この他にも、組織は社内で実行しない機能を実行してもらうためにサービスプロバイダーと契約することがあります。サービスプロバイダーは、自社の顧客である組織に帰属するデータをサービスの成果物として生成します。組織間での内向きデータシェアリングでは、サービスプロバイダーによって生成されたデータが顧客と共有されます。その後、顧客は、自社のデータセンターの外部（ただし、自社のエコシステムの内部）で生成された追加のデータから詳細なインサイトと価値を引き出すためにさらなる分析を実行します。

データを収益化する

今日ではデータは単なる日々のコラボレーションよりも大きな意味を持つ場合があります。データはビジネス資産、つまり通貨です。そのため、データは、データを利用しようとする組織に応じてさまざまな種類の価値をもたらします。したがって、他のあらゆる資産と同様にデータには価値があります。プロバイダーは、データの価値を収益化するためにデータをコンシューマーに販売します。コンシューマーは、そのデータを活用して自社のビジネス目標を推進できます（図 3-3 参照）。

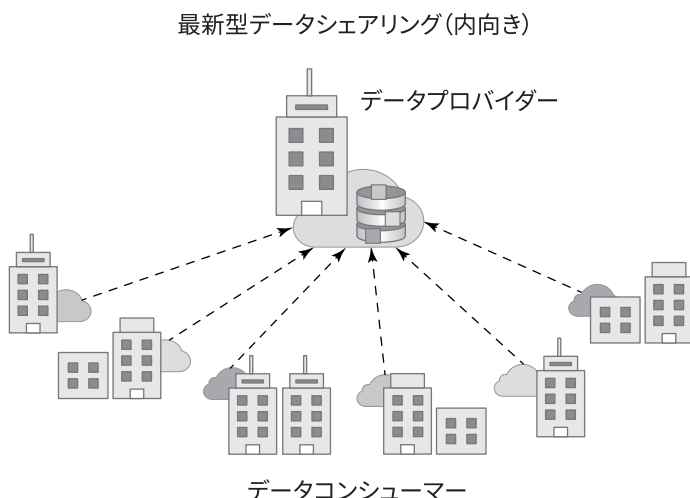


図 3-3: エンタープライズデータプロバイダーは、データコンシューマーとなる他の組織とデータを共有することで新たな収益機会を生み出す。

データコンシューマーは、自社でデータを取得・収集することなしにデータを活用し、そのデータを分析するか、他のデータと組み合わせることで価値を高めることによって直接利益を得ることができます。

しかし、データの利用、大規模なコラボレーション、または付加価値ビジネスの機会のためにデータの価値を活用するには、データを実際に移動することなくデータアクセスを可能にする簡単な方法が必要です。従来のデータシェアリング手法は、コストと手間がかかりすぎる上にリスクが大きすぎます。データを収益化するには、データプロバイダーとデータコンシューマー間でデータを共有するための、セキュアなセルフサービス方式のビジネスモデルを簡単に、かつ低コストで構築できなければなりません。

データシェアリングによって 9,000 万人のゲーマーへのサービス提供を支援



ケーススタディ

PlayFab 社は、月平均 7,500 万人のゲームプレイヤーと日平均 1,500 万人のゲームプレイヤーにサービスを提供しているゲームスタジオの顧客にバックエンドサービスとデータロジスティクスを提供しています。同社は、各ゲームのプレイヤーデータをそれぞれのゲームスタジオと共有しています。また、全スタジオの集約データを匿名化し、その結果をすべてのゲームスタジオの顧客と共有しています。共有データの受領側のゲームスタジオでは、PlayFab 社からのプレイヤーデータを基に、ゲームの最適化を行っています。

PlayFab 社は、データを収集してゲームスタジオと共有する際に、最初の適切なデータの取得、クライアントデバイスからのデータの取得とデータパイプラインへのデータの移動、新しいゲームデータイベントによって常に変化するデータスキーマの管理、大量のゲームデータの移動・転送によって増大するコストの削減などの課題に何度も直面していました。これらの課題を解決するため、PlayFab 社は最新のクラウド構築型データプラットフォームを導入しました。

最新のデータシェアリングにより、PlayFab 社は、シンプルなセルフサービス方式のビジネスモデルの下、各ゲームスタジオとの間でセキュアで管理されたライブのデータビューを設定しています。この手法では ETL はまったく必要ありません。セキュアで管理されたビューにより、各ゲームスタジオのデータが完全に分離され、ゲームデザイナーが、カスタムインポートの必要なしにライブゲームデータのダイレクトフィードに直接アクセスできることが保証されます。

- » 最新型データシェアリングアーキテクチャを探る
- » データコンシューマーのデータベーステーブルとビューへのアクセスを許可する
- » 最新のデータシェアリングの例を確認する
- » セキュアビューで機密データを保護する

第4章 データシェアリングをライブで実現する

本章では、クラウド向けに構築された最新型データプラットフォームアーキテクチャが、データプロバイダーとデータコンシューマーが従来のデータシェアリングの課題を克服するのにどのように役立つかについて説明します。また、最新のクラウドデータプラットフォーム環境内からリアルタイムのデータシェアリングを使用して、データコンシューマー向けのセキュアで管理されたライブデータのビューを迅速かつ簡単に実現する方法についても説明します。

最新型データシェアリングを探る

第2章で説明したように、データシェアリングは多面的な課題です。しかし、従来のデータシェアリング手法では、この課題の一面、つまりデータコンシューマーにプロバイダーのデータへのアクセスを提供することにしか対応できません。従来のデータウェアハウスやデータレイク（分析に必要となるまで大量の生データを保管するリポジトリ）は、データを利用しやすくすることを目的に設計されており、その基盤となるアーキテクチャは、プロバイダーのデータを移動することなくデータコンシューマーにデータアクセスを提供するという最新のデータシェアリングに対応していません。

従来のデータシェアリングは時間がかかり、組織が物事を迅速に実行したい時に足かせとなります。また、従来のデータウェアハウスやデータ

レイクのアーキテクチャでは、特にセキュリティとガバナンスが欠如しているため、図 4-1 に示すように、面倒なデータのアンロードや転送なしにデータコンシューマーによる無制限の同時アクセスやデータプロバイダーによるリアルタイムのデータ変更をサポートすることができません。これにより、データコンシューマーは古い（静的な）データに基づいて運営を行うリスクを負うことになります。

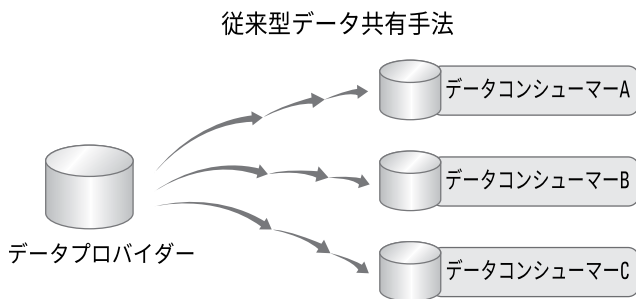


図 4-1: 従来型データシェアリングでは、データプロバイダーとデータコンシューマーの双方が面倒な複数のステップを実行する必要がある。

包括的なソリューションがないため、データプロバイダーとデータコンシューマーはデータを簡単に共有し、データの一貫性を確保するのに苦労しています。また、データを収益化する能力も制限されます。

最新のデータシェアリングで成功を収める

最新のクラウド構築型データプラットフォーム内の最新のデータシェアリングでは、組織（データプロバイダー）は、組織の内外の多数のデータコンシューマーに自社データへのライブアクセスを提供することがわずか数分でできます。社内の事業部間で、またはエコシステム内のビジネスパートナーや社外の組織とデータを共有し、より詳細な分析、データ主導型の取り組み、新たなビジネスモデル、そして新たな収益源を容易にサポートできます（図 4-2 参照）。

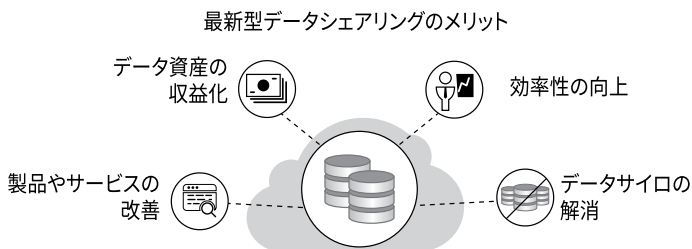


図 4-2: 最新型クラウドデータシェアリングで、データプロバイダーが実現可能なビジネス改善。

最新のデータシェアリングでは、すぐに利用可能なデータがリアルタイムで即座に提供されます。最新のクラウドデータプラットフォームアーキテクチャでは、共有データに対するクエリの速度が飛躍的に向上しており、無限のストレージ・コンピューティングリソースで強化されています。最新のデータシェアリングは、最新のクラウドデータプラットフォームのアーキテクチャと機能をデータの共有に対応できるように拡張します。企業は、セキュアに管理された環境ですぐに利用可能なライブデータ（構造化データと半構造化データ）への読み取り専用アクセスを付与できます。データコンシューマーは、他の組織のデータを結合（JOIN）してデータ分析を強化・深化できます。



ポイント

さまざまなデータソースのデータを保存し、そのデータを多数のデータコンシューマー間でリソースの奪い合いや競り合いなしに共有することを可能にするのは、拡張性、弾力性、柔軟性を兼ね備えたマルチテナントのクラウドデータプラットフォームだけです。

データシェアリングを容易にする

クラウド構築型データプラットフォームは、以下のような機能を提供することで、企業が最新のクラウドデータシェアリングのビジネス上のメリットを実現できるようにします。

- ▶▶ **データの移動やコピーを排除**：最新のデータシェアリングは、管理・制御されたセキュアな環境でライブデータへの直接のリアルタイムのアクセスを提供します。
- ▶▶ **すぐに利用可能なデータを提供**：データコンシューマーは、データウェアハウスの全機能を利用でき、共有データへのアクセスが許可され次第、データのクエリや分析を行うことができます。また、共有データと自社のデータを結合できます。最新のデータプラットフォームは、セキュリティ、ガバナンス、データスキーマ、メタデータをすべて備えています。
- ▶▶ **個人を特定できる情報 (PII) を保護し、業界の要件に準拠**：データプロバイダーは、データを簡単かつ安全に共有し、データコンシューマーによるデータのフル活用を制限することなく、高度な共有機能を使用して PII や機密データの漏洩を防止できます。同様に、データコンシューマーは、データプロバイダーによるデータの強化を目的にデータを共有したい場合には、PII を公開することなく共有できます。
- ▶▶ **追加コストなしにデータシェアリングを実現**：最新のデータシェアリングでは、データコンシューマーがデータをコピーまたは移動することなく、データプロバイダーの共有データを直接閲覧できるため、共有データを保存するためのインフラストラクチャの構築にかかる重複コストを排除できます。

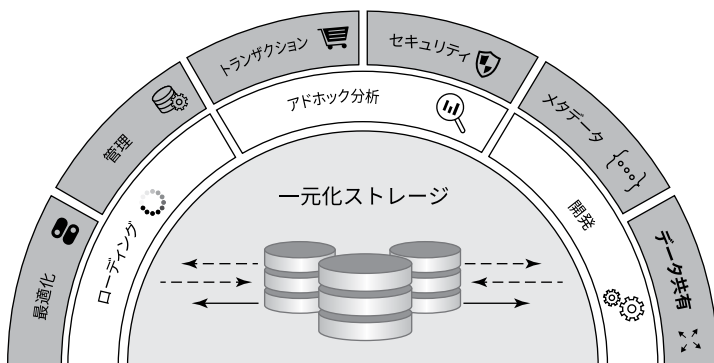
» データプロバイダーとデータコンシューマーの数に制限のないデータシェアリングを実現：最新のクラウドデータプラットフォームは、完全なトランザクションの整合性とデータの一貫性を維持しながら、無制限の数のデータプロバイダーとデータコンシューマーにサービスを提供できます。

最新のクラウドデータシェアリングは、データの公開、アクセス、制御に旧式のメカニズムしか提供できない既存の手法に伴うコスト、遅延、摩擦を解消します。最新のデータシェアリングは、アーキテクチャに関する3つの主要なイノベーションを基盤としています。以下、それらのイノベーションについて説明します。

ストレージ、コンピュートおよびサービスを切り離す

図 4-3 に示すように、ストレージリソースとコンピュートリソースの分離は最新のデータシェアリングのアーキテクチャの根幹部分です。データはすべて、最適化された形式でその忠実性を損なうことなく、クラウドに保存されます。最新のクラウドデータプラットフォームに保存されているデータの単一のコピー（信頼できる唯一の情報源）に任意の数の独立したコンピュートクラスターから同時にアクセスできるため、組織は分析などの社内ワークロードをいくつでも実行できます。

最新型データシェアリングのアーキテクチャ



■ サービス □ コンピュート □ ストレージ

図 4-3: ストレージ、コンピュート、サービスが完全に分離しつつも論理的に統合されている、クラウド向けに構築された最新型データシェアリングアーキテクチャ。

データを共有する場合は、ストレージとコンピュートを切り離すことも重要です。そうすることで、データコンシューマーは自社のデータプラットフォームのコンピュートパワーを使って共有データに直接アクセスで

きます。ただし、(共有データは移動されないため) データコンシューマーはストレージコストを支払う必要はありません。また、データプロバイダーは、データコンシューマーが共有データの分析に使用するコンピュータリソースのコストを支払う必要は一切ありません。

マルチテナントトランザクションを管理する

共有データを利用可能にするには、データへのアクセスと、一貫性、セキュリティ、そしてパフォーマンスを確保するためのすべてのデータコンシューマー間での調整が必要になります。サービスレイヤーは最新のデータシェアリングアーキテクチャの重要な要素です。グローバルなメタデータ、トランザクション、セキュリティはすべてここで管理されます。つまり、サービスレイヤーは、図 4-4 に示すように、データプラットフォームに含まれるすべてのデータベース要素とオブジェクトのデータアクセスを追跡、記録、指示する管制塔の役目を果たします。

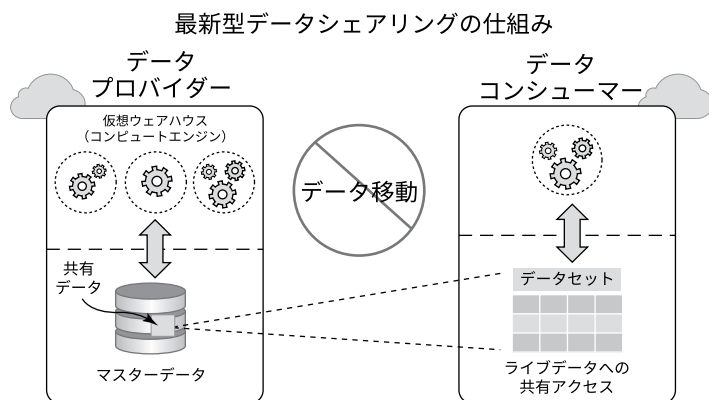


図 4-4: 最新型データシェアリングアーキテクチャのメタデータにより、データ移動なしにデータプロバイダーとデータコンシューマー間のライブのデータアクセスが実現。

さらに、サービスレイヤーは、すべてのデータプロバイダーとコンシューマーにトランザクションの一貫性を提供し、ユーザーに表示されるすべてのデータがライブの最新データの一貫性のあるビューであることを保証します。データプロバイダーは共有データをリアルタイムで更新できます。同様に、すべてのデータコンシューマーは、トランザクションがコミットされた後、データプロバイダーの更新を同時に表示し、共有データに対するクエリを即座に実行できます。これらの操作はいずれも、トランザクションと ACID ベースの一貫性が確保された状態で行うことができます。



技術情報

ACID は、一貫性モデルの一種で、マルチステートメントのトランザクションや処理エラー、さらには電源障害やクラッシュが発生した場合でも、リレーショナルデータベースのトランザクションの有効性を保証する一連の特性を定義します。ACID の特性は以下のとおりです。

- ▶▶ **Atomicity (原子性) (「全か無か」)**: あるトランザクションが完了するには、トランザクションに含まれるすべての処理が成功する必要があります。処理が1つでも失敗すると、トランザクション全体がロールバックされ、データベースの状態は変更されないままになります。
- ▶▶ **Consistency (一貫性)**: 何らかのトランザクションが完了すると、データベースはある有効な状態から別の有効な状態へと遷移します。
- ▶▶ **Isolation (独立性)**: 同時実行されるトランザクションは、データへのアクセスをめぐって競合せず、各トランザクションが順番に実行されるかのように実行されます。
- ▶▶ **Durability (永続性)**: あるトランザクションがコミットされると、そのトランザクションはコミットされたままになります。

データプロバイダーは、コンシューマーが必要とする特定のデータのみを共有することで、プロセスがより容易に、かつ安全になるというメリットも得られます。

制限のない同時実行性を獲得する

最新のデータシェアリングでは、図 4-5 に示すように、多数のデータコンシューマーが共有データに同時にアクセスできます。対照的に、従来型データウェアハウスのアーキテクチャでは、すべてのユーザーがリソースを奪い合うことを余儀なくされるため、最適なパフォーマンスと一貫性を実現することが困難になります。同時実行の自動スケーリングは、追加のコンピューティングエンジンを手動操作なしに自動的に拡張することで、最新のデータシェアリングの同時クエリ処理をさらに進化させます。

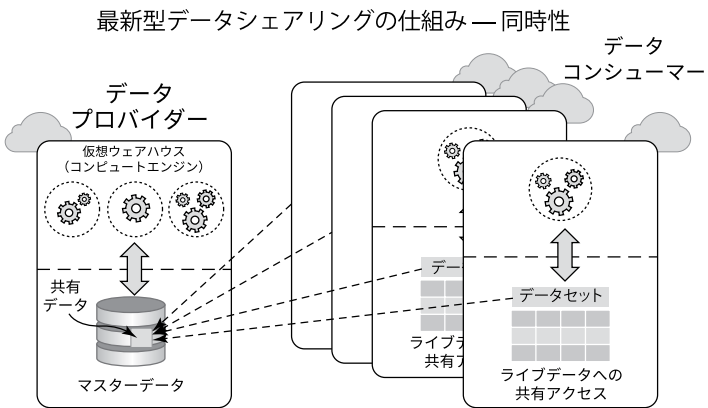


図 4-5: 最新型データシェアリングアーキテクチャによる無制限の同時実行性。

最新型クラウドデータシェアリングを活用する

最新のクラウドデータシェアリングでは、最新のクラウドデータプラットフォームの任意のユーザーがデータベースのテーブルとビューにアクセスできます。データプロバイダーがデータコンシューマーとデータを共有する場合、データベースのオブジェクトとビューはすべて、データプロバイダーの環境内から提供されます。

また、最新のクラウドデータプラットフォームでは、データプロバイダーが、共有を介したデータベーステーブルとセキュアビューへのアクセスをきめ細かく制御できます。データコンシューマーは、アクセス権が付与されている場合にのみ、プロバイダーのデータベースに対してクエリを実行できます。データプロバイダーが共有を作成すると、データコンシューマーがそのデータを照会できるようになります。

データのコピーや移動を伴わない即座のアクセスが可能になるのは、すべてのデータベースオブジェクトが最新のクラウドデータプラットフォームでのみ管理・更新され、そのグローバルメタデータ管理サービスによって調整されるためです。

最新型クラウドデータプラットフォームの仕組みを知る

データプロバイダーがデータを共有するための最初のステップは、特定のデータコンシューマーと共有するデータベーステーブルとビューを指定することです。これはデータ共有オブジェクトを使用して行います。データ共有オブジェクトは、実際のデータベースと共有データベースオブジェクトへの参照を格納する事実上「空のシェル」です。データ共有は、最新型クラウドプラットフォーム環境におけるファーストクラスオブジェクトで、共有を作成・管理するための一連のデータ定義言語（DDL）コマンドを環境に提供します。このコマンドには、`create share`、`alter share`、`drop share`などがあります。アクセスコマンドには、権限の `grant`（付与）と `revoke`（取り消し）があります。共有の作成後、データプロバイダーは、共有する特定のデータベースとデータベースオブジェクトへのアクセスを付与します。SQL セマンティクスは次のようになります。

1. **共有を作成します。** 以下の例では、`sales_s` という名前の空の共有を作成しています。

```
create share sales_s;
```

2. **共有内のオブジェクトに対する権限を追加します。** プライマリオブジェクト内の任意のオブジェクトに対する `USAGE` 権限を

付与するには、先にプライマリオブジェクトに対する USAGE 権限を付与します。たとえば、データベース内のスキーマに対する USAGE 権限を付与するには、先にデータベースに対する USAGE 権限を付与します。データシェアリングコンシューマーを追加する前に、データ共有に対するすべての権限付与を完了します。以下の例では、sales_db データベース、aggregates_eula スキーマ、aggregate_1 テーブルに対する権限を data-share オブジェクトに付与しています。

```
grant usage on database sales_db to
  share sales_s;
grant usage on schema sales_
  db.aggregates_eula to share sales_s;
grant select on table sales_
  db.aggregates_eula.aggregate_1 to
  share sales_s;
```

3. 共有の内容を確認します。

```
show grants to share sales_s;
```

4. 目的のデータコンシューマーに共有へのアクセスを許可します。

以下の例では、データコンシューマー A および B が共有 sales_s を利用できるようにしています。

```
alter share sales_s add
  accounts=data_consumerA, data_
  consumerB;
```

これで data_consumerA と data_consumerB がそれぞれの共有データを表示し、必要に応じて共有データから自分のデータベースを作成できるようになります。

セキュアビューを使用して共有データへのアクセスを制御する

データベースに機密データがある場合はどうなるでしょうか。最新のデータシェアリングでは、共有対象はデータベース全体またはデータベーステーブル全体を共有する必要はありません。テーブルの一部が厳格なセキュリティや機密ポリシーの対象となっている場合、テーブル全体を共有すると、機密データが公開されてしまいます。セキュアビューと呼ばれるコマンドユーティリティを使うと、図 4-6 に示すように、共有データへのアクセスを制御し、セキュリティ違反を回避することができます。

最新型データシェアリングと個別のセキュアビューの連携方法

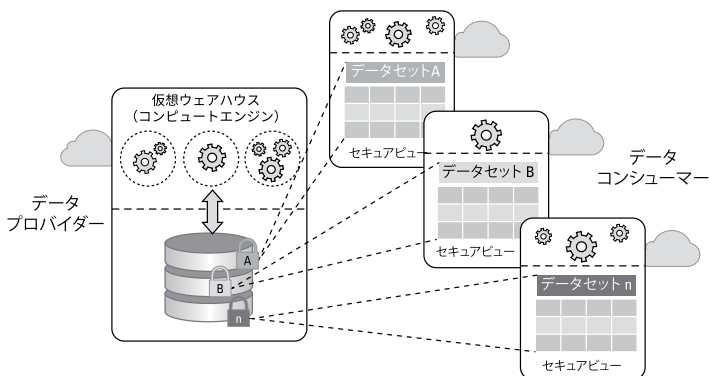


図 4-6:セキュアビューと最新のデータシェアリングにより、データプロバイダーは機密データへのアクセスを保護できる。

たとえば、オンライン小売業者が在庫レベルを計画する場合は、販売業者と商品や販売のデータを共有する必要があります。ただし、販売データを含むデータベース内のテーブルには、機密性の高い顧客 ID 情報も含まれており、この情報をブロックして保護する必要があります。

セキュアビューを使用してこれを達成する方法を示すため、前のセクションで取り上げた SQL セマンティクスに従ってデータベース sales_db のデータ共有オブジェクト sales_s を作成します。この例では、sales_db と sales_s がすでに存在すると想定します。public という sales_db のスキーマにテーブル unitsales が unitsales (customerid, sku, date, qty) として構築されており、そのテーブルにデータが入力されています。

販売業者と新しい在庫を計画する場合、販売台数データへのアクセスを許可する必要がありますが、機密性の高い顧客データへのアクセスは許可したくありません。そのため、unitsales テーブルから販売業者専用の distributor_sales_data というセキュアビューを作成します。

最新型データシェアリングでは、以下のようなステップを実行します。

1. **セキュアビューを作成します。** データベースとスキーマはすでに作成されていて、データが入力されていると想定した場合、次のステップは unitsales テーブルのセキュアビューを作成することです。

```
create secure view sales_db.  
public.distributor_sales_data as  
select sku, date, qty  
from sales_db.public.unitsales;
```

このロジックでは、名前の付いたセキュアビューが作成されますが、機密性の高いcustomerIDデータはこのビューに含まれません。このビューに含まれるデータは、sku、date、qtyです。

2. **sales_s share コンテナで、セキュアビューに対する権限を追加します。**

```
grant usage on database sales_db to
  share sales_s;
grant usage on schema sales_db.public
  to share sales_s;
grant select on view sales_db.public.distributor_
  sales_data to
  share sales_s;
```

このロジックでは、共有（コンテナ）sales_sに、distributor_sales_dataセキュアビューに対する権限が付与されます。

3. **共有の内容を確認します。**

```
desc share sales_s;
```

ユーザーに、共有sales_sの読み出し確認が提供されます。

4. **販売業者に共有へのアクセスを許可します**（ここでは、販売業者も最新型クラウドデータプラットフォームのユーザーであることが前提です）。

```
alter share sales_s add
  accounts=<distributor_name>;
```

5. **共有を確認するには以下のコマンドを使用します。**

```
show shares;
```

この例が示すように、データプロバイダーは、セキュアビューでデータコンシューマーによるデータへのアクセスを制御しながら、データを簡単に共有できます。データコンシューマーは、データをコピーまたは移動することなく、非機密データにアクセスして分析を行うことができます。

本章の内容

- » 真の共有データに基づいてコラボレーションを行う
- » 低コストのコンピュータパワーを活用する
- » 分析を活用してビジネス上の意思決定を迅速にする
- » データの商品化の機会を活用する

第5章 最新型データシェアリングの影響を評価する

本章では、最新型データシェアリングによってどのようにリアルタイムのコラボレーションが実現されるかについて説明します。また、最新型データシェアリングアーキテクチャを可能にしたテクノロジーとトレンド、そして最新のデータシェアリングによって組織がデータから新しいビジネス資産を迅速に作成することがどのように可能になるかについても説明します。

データコラボレーションへの最新のアプローチ

従来、データシェアリングという言葉は、データのコピーを共有することを意味していました（図 5-1 参照）。これにより、以下を含む無数の問題が発生します。

- » 同じデータの複数の静的で同期されていないバージョンがさまざまな環境で複数のデータサイロにまたがって存在する。
- » 組織内のデータに実用的な信頼できる唯一の情報源またはガバナンスが存在しない。

- ▶▶ 古い、不完全、あるいは不正確なデータを基に重要なビジネス上の意思決定が下される。
- ▶▶ 訴訟支援のために組織内外の複数のデータソースを特定、検索、生成しなければならない場合に電子証拠開示コストが増大する。
- ▶▶ 潜在的なデータ侵害件数や、偶発的データ損失・開示のリスクが増大し、侵害通知、信用監視サービス、企業ブランドへのダメージ、顧客離れ、訴訟、フォレンジック分析、復旧といった関連コストが膨れ上がる。

従来型データシェアリングの問題点

プロバイダーと消費者の双方が問題を抱える

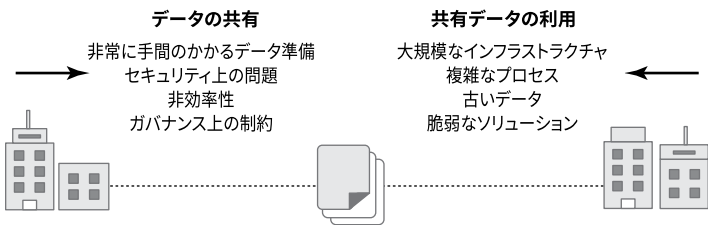


図 5-1: 従来型データシェアリング手法の難点。

最新型データシェアリングでは、データ消費者が同一データの単一コピーにリアルタイムでアクセスできるため、迅速かつ費用対効果の高い安全なコラボレーションが組織の内外で実現します。また、データプロバイダーは、組織の運営に使用しているものと同じ最新のクラウドデータプラットフォームを使用してデータを共有できます。

コンピューティングの民主化

エンタープライズグレードのコンピューティングパワーはあらゆる場所にあり、日々よりパワフルに、より高速に、そしてより安価になっています。クラウドによってこのトレンドがさらに促進され、あらゆる規模の組織が無数のコンピューティングリソースを大規模かつ容易に利用できるようになりました。

この巨大なパワーにより、ビジネスユーザーはさまざまな分析タスクを自分で実行できます。IT 管理者にジョブリクエストを送り、実行してもらうまで待つ必要はもうありません。ビジネスアナリストであれば、大規模なデータセットに対して高度な分析を実行する力が手に入ります。データサイエンティストであれば、予測分析を行い、人工知能の基礎となる機械学習アルゴリズムを開発できます。経営幹部であれば、最新の分析ダッシュボードを駆使して迅速に意思決定できます。製品管理担当幹部であれば、新しい製品・サービスの市場投入にかかる時間を短縮できます。簡単に言うと、今や誰もが巨大なコンピューティングパワーを活用できるということです。

データ主導のビジネスの意思決定

今日では組織は、顧客の全体像を把握するため、マルチチャネルアプローチを採用し、さまざまなチャネル（Web サイト、モバイルデバイス、POS 端末、コールセンター）からデータを収集する必要があります。従来型データウェアハウスは、データを共有できないだけでなく、さまざまな形式の構造化データと半構造化データの間の有意義な関係を明らかにする能力も十分ではありません。



技術情報

JSON などの半構造化データは、従来の構造化データの標準に準拠していません。また、このようなデータには、ソーシャルメディアサイト、クリックストリーム、モバイルデバイス、モノのインターネット（IoT）デバイスなどの新しいデータソースによって生成されたデータが含まれています。

しかし、組織が成功を収めるためには、このようなデータを複数のチャネルから収集して分析することが欠かせません。インサイトに満ちた分析から得られた結果は、組織が新しい製品やサービスのターゲットをより適切に設定するのに役立ちます。そのため、データをサービスとして、または付加価値のあるビジネス資産として提供することを望んでいる組織が、競合しない他の組織にとって利益になるようにデータへの迅速かつ容易なアクセスを提供することに関心を持つのも当然のことです。

コストを削減し、運用を効率化



ケーススタディ

Heap Analytics 社は、さまざまな業界の企業に Web やモバイルの分析サービスを提供しています。以前まで同社は、顧客と生データを共有するため、ファイルを Amazon S3 バケットにドロップするか、後でデータを分析するために必要な ETL（抽出、変換、ロード）作業に完全に対応できるほど高度な能力を持たない顧客向けにクラウドベースのベンダーと協力してファイルをホストしていました。当時、同社が負担していたコストと実行していた作業は以下のとおりです。

エンジニアリング作業：パイプラインのデバッグ、クラスターのメンテナンス、顧客ごとに異なる設定への対応、リソースの競合の解決に週 5 ～ 10 時間

カスタマーサービス作業：適切なアクセスの確保、クラスターの運用のスケジューリング、従来型クラウドデータウェアハウスにホストされたクラスターの作成、メンテナンス、サイズ変更に関し週 5 時間以上のソリューション時間

Heap 社は、最新のセキュアなデータシェアリング機能を備えたクラウド構築型データプラットフォームを活用することで、エンジニアリング作業をなくし、カスタマーサービス作業を大幅に削減できました。また、他の分野でもメリットが得られました。たとえば、顧客とのデータ共有が可能になり、顧客は即座にデータへアクセスして価値実現までの時間を短縮できるようになりました。Heap 社は、最新のソリューションの導入により、権限とセキュリティの承認に関連する繰り返し作業の減少、カスタム設定の排除、インフラストラクチャのばらつき軽減、クラスターのメンテナンスの容易化、リソースの競合の解消を実現し、サポートの負担を減らすことに成功しました。

データの商業化

最新のデータシェアリングは、オンプレミスおよびクラウドデータウェアハウスの従来の障壁を取り除くことで、少なくとも以下の 4 つの経済的機会を企業にもたらします。企業は、これらのチャンスを活用することで、データをパッケージ化・収益化された資産として迅速かつ安全に共有し、真のデータエコノミーに活力を与えることができます。

- » **データの収益化：**多くの企業がデータを生成して販売しています。このようなデータの収益化が始まったのは 100 年近く前です。現在は、最新のデータシェアリングにより、どのような組織でも、データリポジトリのスライスへのアクセスを課金対

象とすることで、任意のサイズのデータをビジネス資産に変えることができます。この低コストで悩みの種がないソリューションにより、データ企業は、データの価値を最高水準に保つ最新の精度を備えた新鮮なデータに対するデータ消費者の差し迫った要求に即座に応えることができます。組織は、データ収益化戦略をサポート、推進、実行するための大きなパワーとプラットフォーム機能が手に入ります。

- ▶ **ビジネスパートナーとのデータシェアリング**：ビジネスパートナーとデータを直接共有することは目新しいことではありません。しかし、ライブデータを容易に共有できることは画期的なことです。最新のクラウドデータシェアリングを導入すると、以下のことができるようになります。
 - 組織はエコシステムのビジネスパートナー（サプライチェーン、流通、マーケティング、サードパーティ販売など）と即座にデータを共有できます。
 - どちらの組織がデータを所有しているか、データプロバイダーとデータ消費者の役割をそれぞれどちらの組織が担うかに関係なく、データ消費者がライブデータに対してクエリを実行できるため、データ消費者とデータプロバイダーの双方に利益がもたらされます。
 - 分析の基盤となる内部および外部のデータに、即座に、オーバーヘッドなしで、従来のデータシェアリング手法の数分の1のコストでアクセスできるため、分析の実行が迅速化され、ビジネス成果がより迅速に実現されます。
- ▶ **エンタープライズデータサイロの解消**：最新のデータシェアリングは、構造化データと半構造化データを最新のクラウドデータプラットフォームに保存できるようにすることで、組織の内外に散在するデータサイロを解消します。ダウンロードや複製を行うことなくシームレスに共有できます。手作業による統合やデータパイプラインの必要なしに、これまでサイロ化されていたシステムを緊密に統合できます。
- ▶ **管理の手間がほとんどかからず、より多くのデータ消費者にデータを提供することが可能**：旧式のデータウェアハウスの管理に必要な従来型の時間のかかる手法を排除できるのは最新型クラウドデータプラットフォームだけです。最新のデータプラットフォームはパフォーマンスを念頭に置いて設計されているため、インフラストラクチャを調整することも、ノブを回すことも、チューニングすることも不要です。管理の必要がほとんどないため、組織はより広範なデータシェアリング戦略を推進し、組織内の事業部門、エコシステム内のビジネスパートナー、デジタルエコノミーを構成するその他の組織を含むより大きなデータ消費者ベースをターゲットにできます。

数千のデータコンシューマーと信頼できる唯一の情報源



ケーススタディ

PlacelQ社は、モバイルデバイス上の何千というアプリケーションからのデータを集約、収集、匿名化し、モバイルコンシューマーの位置情報や行動を基にターゲットを絞り、そのターゲットにアプローチしたい企業にそのデータを提供しています。PlacelQ社の顧客には、マーケティング会社、広告代理店、製品メーカーなどがあります。

1つのデータソース - 数千のセルフサービスユーザー：PlacelQ社の顧客は、モバイルデバイスユーザーの個別の使用状況と地理的移動に基づいて高度にターゲットを絞り込んだキャンペーンを展開しています。これを行うため、各顧客は潜在的な見込み客を特定するために個別のデータスライスが必要とする場合があります。従来手法はコストと手間がかかりすぎるため、セルフサービス方式のビジネスモデル内でデータを容易に配信することができませんでした。

効果的なビジネスモデル - ゼロデータ管理：データシェアリング向けに構築された最新のクラウドデータプラットフォームにより、何千もの個別のデータコンシューマーサブスクリプションに拡張し、管理されたセキュアなデータスライスを各コンシューマーに提供するというPlacelQ社が直面していた課題が解決されました。PlacelQ社は現在、信頼できる唯一の情報源の管理に使用するリソースを大幅に減らしながら、データコンシューマーが自社のデータサブスクリプションをセルフサービスで管理できるようにしています。PlacelQ社は、最新のデータシェアリングを活用し、顧客ごとに差別化され、それに応じて価格設定されるデータの新しいユースケースを開発することに重点的に取り組んでいます。

PlacelQ社のデータコンシューマーにとってのシンプルさ：PlacelQ社の顧客は、共有データを自社のデータセットと統合する前にシンプルな環境で共有データを評価できるというメリットも得られます。これを行う際に、必要なデータを見つけるために大量の無関係なデータを取り除く必要はありません。さらに、最新型データプラットフォームから得られるデータはすべて、一般的なセキュリティと認証の要件を満たすため、顧客はコンプライアンス上の問題に煩わされることなくデータセットを統合できます。

本章の内容

- » データエクステンションを作成してデータをサードパーティに拡張する
- » リアルタイムのデータシェアリングを設定する
- » 従来のデータマーケットプレイスの課題を克服する

第6章

エクステンションを活用してデータを共有し、インサイトを収益化する

制御されたエコシステム内でデータを交換すると、豊富な分析結果と深いインサイトが得られ、より多くの情報に基づいて意思決定を下すことができます。本章では、データエクステンションに参加し、サードパーティのデータを活用して新しいデータ資産を作り出す方法について説明します。

データマーケットプレイスとデータエクステンションの歴史を検討する

データマーケットプレイス（データエクステンションとも呼ばれます）では、利用者が有用なデータセットを見つけることができます。従来のデータエクステンションには、利用者はポータルやアプリストアのような環境を通じてアクセスします。多くのデータプロバイダーは、基本的なデータセットを無料で提供し、追加のデータセットやデータサービスに料金を課しています。

今日では、何千というオンライン型マーケットプレイスが売り手と買い手を結びつけています。通常は、データベンダーがデータの変換、準備、読み込みを行い、マーケットプレイスが検出、コラボレーション、ライセンス、監査を担当します。データマーケットプレイスは非常に大きなチャンスです。Accenture は、2018 年 9 月発表のレポート「データの価値：データマーケットプレイスの夜明け」で、サービスとしてのデータ (DaaS) の市場規模は 2021 年には 104 億ドルに達すると予測しています。

これに対し、最新型データシェアリングのアーキテクチャを基に構築された最新のデータエクステンジでは、データプロバイダーとデータコンシューマーの間で、1 対 1、1 対多、そして今では多対多のデータシェアリング関係を確立できます。また、ほとんどのデータマーケットプレイスで悩みの種である従来のデータ転送に伴うコストと手間が排除されます。

ほとんどの組織は、データシェアリングとクラウド構築型データプラットフォームを使用してすでにデータを共有しているか、計画しています。データを収益化できるマーケットプレイスが存在しており、その規模は巨大で急成長しています。データコンシューマーとデータスライスを共有することで収益を増やせるだけでなく、自社のデータとその他のプロバイダーのデータおよびデータサービスを組み合わせることで新たに提携を結んだり、新しいインサイトを見つけ出したりすることもできます。データの収益化については、第 7 章で、詳しく説明します。

データエクステンジを利用する理由

エクステンジを使用して共有データを提供・利用することで以下のことを実現できます。

- **マーケットのチャンス創出**：データセットの読み取り専用バージョンへのセキュアな管理されたアクセスをデータコンシューマーに有料で提供することで、新たな収益源を開拓できます。
- **セキュリティとデータアクセスの改善**：読み取り専用データセットへのセキュアなリアルタイムの接続を開くことで、データのコピーに伴うリスクとコストを最小限に抑えることができます。
- **カスタマーエクスペリエンスの向上**：データとデータサービスにアクセスするための高速かつ安全で費用対効果の高い方法を顧客に提供することで、データの利用率を増やし、データサブスクライバーの満足度を高めることができます。

リアルタイムのデータシェアリングで最新化を推進する

ほとんどの企業は、データエクステンジからかなり迅速に価値を引き出しており、追加のデータとデータサービスから得られる詳細なインサイトを活用しています。このようなデータコンシューマーは、APIや探索・分析用のセルフサービスインターフェイスを使用する、ダウンロードやファイル転送によって生データを購入する、トレンドやインサイトを明らかにするアプリを活用するなどの複数の方法でサードパーティのプロバイダーからデータを取得しています。その後、次の2つの一般的な疑問が生じます。

- ▶▶ 顧客やパートナーなどの社外関係者を含めるようにデータシェアリングの取り組みを拡大するにはどうすればよいか？
- ▶▶ 分析業務を圧迫することなく、サードパーティから簡単に入手できる関連データを見つけるにはどうすればよいか？

この2つの疑問に対する答えは、アーキテクチャによって異なります。最新のデータエクステンジでは、コンシューマーとのデータ共有に、データの移動も、データの抽出、変換、ロード（ETL）テクノロジーも、継続的な更新作も必要ありません。ファイル転送プロトコル（FTP）を介したデータ転送も、APIの開発・保守も必要ありません。また、データはコピーされるのではなく共有されるため、追加のクラウドストレージも不要です。データプロバイダーは、図6-1に示すように、簡単にデータを公開し、データコンシューマーがすぐにデータを発見、照会、そして強化することができるようにすることができます。

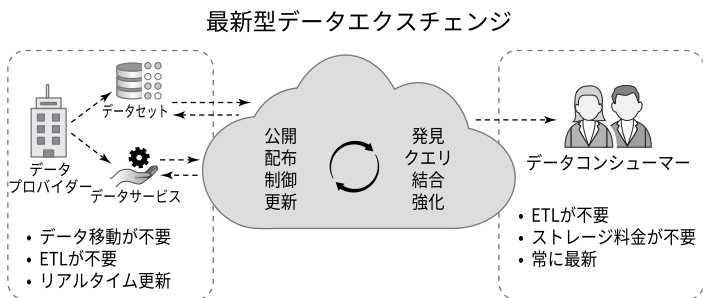


図 6-1: 効率的なリアルタイムデータエクステンジを実現するアーキテクチャ。

データエクステンションに適したアーキテクチャを構築する

APIの開発・保守は、データプロバイダーとデータコンシューマーにとって大きな負担になります。プロバイダーはAPIを設定・管理する必要があり、コンシューマーは取引相手となるすべてのマーケットプレイスベンダーへのカスタムリンクを確立する必要があります。その代わりに最適なデータエクステンションは、プロバイダーとコンシューマーの双方に以下のメカニズムを提供します。

- ▶ APIやETLプロセスを使用せずにライブデータセットにリアルタイムで容易にアクセスできる。
- ▶ データを移動または変換することなしに構造化データセットと半構造化データセットをアップロードできる。
- ▶ データのさまざまなビューを作成し、個人を特定できる情報(PII)を保護しながらそれらのビューを安全に共有できる。
- ▶ すぐに利用可能な形式のデータをデータコンシューマーに提供できる。

一部のクラウドベンダーは、マルチテナントSaaSアーキテクチャの固有の機能を活用して顧客にデータエクステンションオプションを提供しています。主要クラウドベンダーは、このようなエクステンションで良好なパフォーマンスを確保するため、コンピュートリソースとストレージリソースを分離したマルチクラスターの共有データアーキテクチャを提供する必要があります。このアーキテクチャにより、データプロバイダーはコストをほとんどかけずにデータを共有できます。一方、データコンシューマーはコンピュートリソースの料金を払うだけでデータを分析できます。ストレージリソースの料金を支払う必要はありません。

利便性を最大限に高めるため、エクステンションは登録無料で、利用者がデータセットを即座に公開でき、エコシステムの他の承認済ユーザーがいつでも必ず最新のデータにアクセスできるものでなければなりません。データストレージレイヤーは、Amazon Simple Storage Service (S3)、Microsoft Azure、Google Cloud Platform (GCP)などのスケラブルな従量課金制のクラウドレイヤーに設置する必要があります。データプラットフォームプロバイダーは、すべてのものが安全に保たれ、セルフサービスアクセスのために適切に管理、調整、最適化されるように、これらのストレージ資産をサービスレイヤーでラップします。このタイプのアーキテクチャにより、動的な弾力性を備えた単一の集中型システムからデータを効率的に交換することが可能になります。データプロバイダーは無制限の量のデータを共有でき、受信者は実際に使用したデータとリソースの料金のみを支払います。

最適なデータマーケットプレイスを構築する



ケーススタディ

気象データは、あらゆる業界のビジネスインテリジェンスにとって重要かつ不可欠な資産です。気象による収益の減少額は米国だけで年間600億ドル以上に上ります。気象戦略は、企業による売上の予測、リスクの軽減、輸送ルート調整、確信のある意思決定に役立ちます。

Weather Source社のチームには、75年間にわたって蓄積してきた気象学と気候学の知識があります。同社は、2000年から現在までの気象データと最大15日間の気象予報を提供しています。同社のソリューションは超地域密着型であり、顧客のビジネスにとって最も関連性の高い地点に調整できます。

Weather Source社は、そのデータ配信サービスへのアクセスを拡大するため、データエクスチェンジサービスに参加しました。セキュアなデータシェアリングプラットフォームを基盤に構築されたデータエクスチェンジは、データに簡単にアクセスするための手段を企業に提供します。

データエクスチェンジにより、Weather Source社はデータセットの公開にかかるコストと労力を削減できるようになりました。データの変換、ロード、再構築はもう必要ありません。新しいオブジェクトが公開されると、すべてのコンシューマーがすぐにそれを利用できるようになるため、エコシステム全体でリアルタイムのデータアクセスが実現します。この効率的なオペレーションにより、顧客はわかりやすいSQL形式で常に最新のWeather Sourceデータにより便利にアクセスできます。



ヒント

長期的な柔軟性を備え、市場が拡大し続けるデータエクスチェンジに参加したい場合は、複数の地域へのデータの複製とAWS、Microsoft Azure、Google Cloudなどのさまざまなクラウドへのデータの保存をシームレスなエクスチェンジの一環としてサポートするクラウドデータプラットフォームプロバイダーを選ぶとよいでしょう。そうすれば、データエクスチェンジが進化したときに1種類のオブジェクトストアや特定のクラウドベンダーに閉じ込められることはありません。

一部のクラウドデータプラットフォームプロバイダーは、顧客が自己完結型のエコシステム内でプライベートデータエクスチェンジを作成し、データとデータサービスを共有・収益化することを許可しています。データプラットフォームの顧客は、このエクスチェンジを活用し、最終顧客とパートナーにデータとデータサービスを提供することで新しい事業を開拓できます。コンシューマーは、これらのデータエクスチェンジに容易にアクセスでき、プロバイダーが定めた条件に従って、利用したデータとデータサービスの料金のみを支払います。

データエクステンションベンダーによる支援を受ける



ヒント

最新のデータエクステンションでは、以下のようなさまざまな固有のメリットが提供されます。

- さまざまな業界と機能にまたがる何千ものデータセットにアクセスできる
- データを検索して発見し、詳しく分析することが簡単にできる
- 手動操作なしでライブデータセットにアクセスできる
- きめ細かい履歴データに完全にアクセスできる
- ETLプロセスを構築して維持する必要がない
- データを移動またはコピーする必要がないため、追加のデータストレージコストが発生しない
- 付加価値の高いデータサービスにアクセスしてデータ分析を強化できる



ポイント

従来のエクステンションは、FTP ダウンロード、ETL 変換、API 接続などに依存しています。最新型のエクステンションは、ストリーミングデータサービスや SQL クエリを使用して情報を抽出します。データの変換やコピーは必要ありません。

- » データ収益化の機会を見つける
- » ガバナンスモデルと流通モデルを評価する
- » DaaS（サービスとしてのデータ）の価格を設定する

第7章 データを収益化する

データの収益化とは、データプロバイダーがデータコンシューマーに自社のデータやデータサービスへのアクセスを有料で提供し、データコンシューマーが既存のデータセットを拡充してそのビジネスと顧客にメリットをもたらせるようにするプロセスです。データそのものへのアクセスを提供するデータプロバイダーもあれば、データモデリング、データエンリッチメント、データ分析などのサービスを提供するプロバイダーもあります。

Nielsen 社はこの分野のパイオニアで、そのサービスは広範に普及しており、その名は広く知られています。同社は、90年以上にわたり、消費者データ収集および分析し、メディア企業、広告会社、小売企業などのさまざまな業界の組織に販売しています。Nielsen 社の実績あるビジネスモデルは、これまでニッチなデータシェアリングの機会を数多く生み出してきました。たとえば、金融サービス業界では、いくつかの企業が株式市場データを収集し、それをパッケージ化して証券会社やヘッジファンドに販売しています。ゲーム業界では、PlayFab 社がライブゲームスタジオの顧客（コンシューマー）にゲームと分析サービスを提供し、10億人以上のゲームプレイヤーが利用している 2,500 以上のゲームを強化していることで知られています。PlayFab 社は、プレイヤーのアクティビティに関するリアルタイム分析、ダッシュボード、カスタムレポートをデータコンシューマーに提供しています。これにより、スタジオはプレイヤーのエクスペリエンスを最適化し、ビジネスをさらに推進することが可能になります。

業界の機会を検討する

図 7-1 に示すように、組織は無数の方法でデータを収益化できます。たとえば、電気通信事業者であれば、小売業者が広告で消費者をターゲティングするのに役立つ位置情報データを販売できます。消費財メーカーであれば、購買データをオンライン広告主と共有したり、顧客と直接共有したりできます。物流企業であれば、輸送パターンや海運活動に関するデータを経済動向の指標として販売できます。輸送機関であれば、料金所や橋の通行データを収集し、ナビゲーションアプリを最適化して交通渋滞を最小限に抑えることができます。データシェアリングにより、エンジンパフォーマンスの追跡からヒューマンバイオメトリクスのモニタリングまで、これまで入手できなかったインサイトが明らかになるため、組織はトレンドを特定したり、予測を行ったり、是正措置を講じたりすることが可能になります。

データへのアクセスはさまざまな業界で新しい価値を生み出す



図 7-1: あらゆる業界で貴重なチャンスの扉を開くデータシェアリングデータへのアクセスにより、さまざまな業界で新しい価値のチャンスが開く。

同様に他の組織からデータを取得することで、社内の業務を強化できます。たとえば、カスタマーサービス部門であれば、見込み客を分類したり、豊富な知識を持ってチャンスを転送したりするために発信者に関する外部データが必要になる場合があります。マーケティング担当者であれば、メッセージの対象を絞り込んだり、キャンペーンを調整したりするために外部のデータソースに目を向けます。営業部門であれば、販売先の組織に関するサードパーティのデータを収集・集計し、リードを適切に選別できます。リスクアナリストであれば、サイバーセキュリティの専門家からデータを収集し、ネットワーク侵入者を特定したり、不正行為を防止したりするために活用できます。

データ収益化のトレンドに正面から取り組む

組織が収集するデータ量の飛躍的増加に伴い、最新型のデータシェアリングは年々成長しています。Forrester Research 社がグローバルデータと分析の意思決定担当者を対象に行った調査によると、75%以上の企業が社外データの利用を拡大していると回答し、それらの企業の約50%がすでにデータの商業化を行っています。

インターネットトラフィックから気象データ、ソーシャルメディアのトレンド、購買パターンまで、貴社のデータはサードパーティにとって価値がある可能性があります。貴社がデータを収益化しているかどうかにかかわらず、競合他社がすでに着手している可能性があります。IDCは、「2019 Predictions for Digital Transformation (2019年のデジタルトランスフォーメーションの予測)」において、80%の企業が2020年までにデータの管理と収益化の機能を開発し、2023年までに95%の事業者が新しいデジタル KPI セットを導入すると予測しています。

チャンスがどこにあるかを見極めるには、次のような質問を自分自身に問いかけてください。

- ▶▶ 顧客はこれまでアクセスできなかった特定のデータへのアクセスを要求しているか？
- ▶▶ 顧客はより多くのデータ、より良いデータ、より良いインサイト、または新しいインサイトを求めているか？
- ▶▶ 顧客は現在のデータ共有方法に不満を感じていないか？顧客の要望に応えることは社内スタッフにとって重荷になるか？



ヒント

まず、顧客にとって価値のある分野を特定してください。分野が特定できれば、顧客に提供するデータとアクセスのタイプと、将来のデータシェアリングの機会を最大化するために現在のビジネスのやり方を拡張する方法がわかります。

データの収益化戦略を立てる際には、以下のオプションを検討してください。

- ▶▶ 生データへのアクセスを販売する
- ▶▶ 特定のデータサービス（データエンリッチメント、データモデリングなど）を有料にする
- ▶▶ データ分析を有料で提供する

優れたデータガバナンスの遵守

図7-2に示すように、あらゆるタイプのデータシェアリングのシナリオにおいて、コンシューマーデータのセキュリティとプライバシーに関する政府および業界のすべての規制を遵守する必要があります。共有するデータがどのようなものにかかわらず、条件や制限事項を設けておくことが重要になります。これらの法的指令に注意を払うデータガバナンス責任者を任命し、社内のポリシーと手順を定め、顧客の要求に応え、セキュアなビュー、セキュアな関数、セキュアな結合を使用してデータへのアクセスを制御してください。これらについては、次章で説明します。

データガバナンスの課題

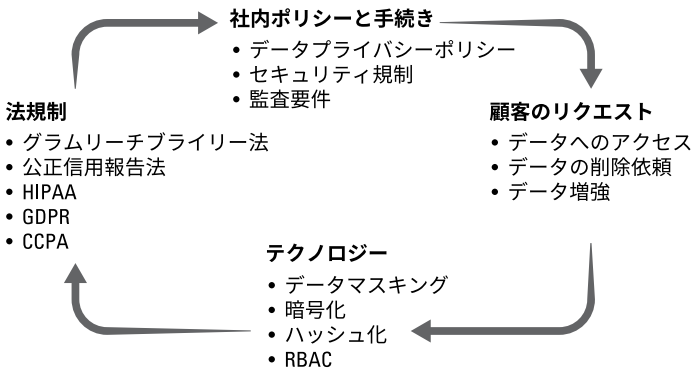


図 7-2: データシェアリングのやり方が関連するすべての規制に準拠していることと顧客のプライバシーを侵害しないことを確認する。

販売チャネルを評価する

本書の他の箇所でも説明しているように、データを共有するには、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の交換を簡素化する安全で柔軟なビジネスモデルが必要です。データとデータサービスを販売し、市場に勢いを与えるための流通チャネルとして以下が考えられます。

- » **直接顧客へ:** このアプローチのメリットは、よく知られているオーディエンスに販売し、それらのオーディエンスとの関係を深める機会が得られることです。デメリットは、市場が限定されることと専用のシェアリングプラットフォームの構築が必要になることです。

- ▶▶ **データブローカー経由**：メリットは、データブローカーが既存顧客を紹介してくれるため、導入が容易になることです。デメリットは、データをコントロールできないこととユーザーと直接的な関係を築けないことです。
- ▶▶ **データエクステンション経由**：第6章で説明したように、データエクステンションは柔軟で双方向のデータシェアリングをサポートするスケーラブルなプラットフォームです。これはブランドを構築し、顧客との新しいタイプの直接関係を築くための優れた方法です。考えられる唯一のデメリットは導入に関するものです。初期段階のエコシステムから開始する場合は、サービスがその潜在力を最大限に発揮するまでに、時間がかかることがあります。このような場合には、誰かが見つけるまでデータをそのまま置いておく従来型のデータマーケットプレイスではなく、価値提案の一環としてブランド認知度の向上と規模の拡大に寄与できるデータエクステンションを探してください。

データシェアリングサービスの価格を設定する

顧客やパートナーにデータの料金を設定する方法はいくつかあります。ストリーム配信されるデータや継続的に更新されるデータの場合は、月額サブスクリプション料金を設定するのが一般的です。ただし、データセットが比較的静的である場合は、初回利用時に価値の大部分が得られ、時間の経過とともに価値が低下します。このような場合には、顧客にサブスクリプション料金を請求するよりも1回限りの料金を支払うよう求めるほうがコスト効率が高くなる可能性があります。

また、以下のようにさまざまなサービスレベルに応じて段階的に価格を設定することも検討できます。

- ▶▶ **無料**：特定のデータセットへの限定アクセス
- ▶▶ **標準**：データセットへのフルアクセス
- ▶▶ **プレミアム**：データセットへのフルアクセスに加えてモバイルアラート、基本的なインサイト、カスタムダッシュボード、ロールベースのレポートを提供
- ▶▶ **エリート**：プレミアムの特典に加えて予測分析と意思決定支援ツールを提供

データシェアリングサービスが切り開く新たな収益源



ケーススタディ

FMCG（Fast Moving Consumer Goods、日用消費財）業界で製品品質を維持するには、在庫、倉庫管理、出荷、配送に関する迅速な意思決定が求められます。Atheon Analytics社は、食料品卸売業者がデータのトレンドを可視化し、小売業者とインサイトを共有するのを支援することで、商品を適切な場所に適切なタイミングで移動し続けています。同社のSKUtrak サービスは、英国全土のFMCG サプライヤーにとって主要な商品フロートラッカーです。

新鮮な食料品は新鮮なデータの副産物です。Atheon社は、同社のクラウド構築型データプラットフォームを基盤とする新しいデータシェアリングサービスを活用し、SKUtrakの顧客と簡単にデータを共有しています。これにより、顧客は複雑なデータコピー手順やデータ移動手順を踏むことなく、常に最新のインサイトを閲覧できます。Atheon社のサービス提供責任者であるRose Ahearne氏は次のように述べています。「当社は、小売業者から直接データを提供し、それをサプライヤーやお客様と共有することで、商品を必要としているスーパーマーケットに商品が確実に届くようにしています。このデータプラットフォームのおかげで、スケーラビリティを確保し、新商品開発の機会を生み出すことができました」。

リアルタイムのデータを安全に共有できるクラウド構築型の分析プラットフォームは、今や同社のビジネスに欠かせない要素となっています。SKUtrakではデータを可視化することにより、小売業者とそのサプライヤーは通常のパターンと異常がある点を調べ、食料品サプライチェーン全体で改善のチャンスを見つけることができます。その結果、供給体制が改善され、無駄が減り、買い物客のニーズにより良く応えられるようになりました。Ahearne氏は次のように付け加えています。「データシェアリングは素晴らしい収益源を新たに開き、当社の製品ポートフォリオの一部になりました。開発者とコンサルタントはデータを待つことがなくなり、生産性が大幅に向上しました」。

- » 政府と業界の規制を遵守する
- » 機密データを安全に保つ
- » データセットの共有とアクセス制御を選択的に行う

第8章 データを統制する

多くの組織は、自社のデータとパートナー、サプライヤー、顧客、同業他社によって生成されたデータを安全に結び付ける必要がありますが、個人を特定できる情報 (PII)、保護医療情報 (PHI)、競合データ、その他の形式の高粒度データの保護に懸念を抱えています。

これは喫緊の課題です。今日の企業は、EU の一般データ保護規則 (GDPR)、米国で 1996 年に制定された医療保険の相互運用性と説明責任に関する法律 (HIPAA)、カリフォルニア州消費者プライバシー法 (CCPA) など、消費者データのセキュリティとプライバシーを規定する厳格な規制に従う必要があります。これらの規制は、データの作成と保存から利用と共有、アーカイブと削除まで、データのライフサイクル全体を通じて遵守する必要があります。

GDPR を例にとると、考慮すべき重要ポイントは以下の 3 つです。

- » 個人データの処理方法と取り扱い方法に関する原則
- » 個人 (データ主体) の権利
- » 説明責任 (データのコンプライアンスを文書化および実証できることを含む)

GDPRにより、組織は複数のデータソースによって生じる複雑さの増大を抑制するとともに、世界的規模で最新のデータガバナンス戦略を確立することが求められます。ほとんどの組織は、複数の場所（異種のデータウェアハウス、データマート、サーバー、クラウド）にデータを保存しているだけでなく、隠れたデータレイヤーに悩まされてもいます。隠れたデータレイヤーとは、サイロ化したチームと部門による情報共有の結果として組織内に無数に存在するデータコピーのことです。

データを共有するためにデータをエクスポートしてコピーする必要がある従来のオンプレミスデータハウスのような旧式のテクノロジーに依存すると、データコピーがまん延します。また、クラウドストレージ、ファイル転送プロトコル（FTP）、APIを経由したデータセットの配布が追跡困難であるなら、社員がサムドライブやメールを介してファイルやスプレッドシートを手動で共有する場合はどうでしょうか？適切なテクノロジーがない場合、コンプライアンス要件を満たすのは困難です。なぜなら、データの一部または全部がコピーされる場合、おそらくは複数の場所にある複数のデータセットに対応しなければならなくなるからです。

幸いなことに、データを共有するための適切なテクノロジーを導入すれば、データのプライバシーと保護の要件を満たすことが大幅に簡単になります。多数の場所にある多数のデータに対応する代わりに、高度なテクノロジーによってアクセスと共有が管理される単一のデータコピーを確立できます。



ポイント

履歴データとライブデータの唯一の「信頼できる情報源」を1つの場所に保管し、そのデータコピーの管理されたスライスへのオンデマンドのアクセスを許可することで、コンプライアンスの取り組みを容易にすることができます。

必要なデータセキュリティのタイプを判断するには、最初に自社のユースケースに最も近いデータシェアリングのシナリオを特定します。本章では、そのようなシナリオを5つ取り上げ、それらについて説明します。

同じデータを1つまたは複数のコンシューマーと共有する

この状況では、必要な作業はシンプルです。1つまたは複数のデータコンシューマーと共有するテーブルまたはビューを用意します（図8-1参照）。たとえば、小売チェーンが50のフランチャイズとデータベース全体を共有したい場合は、この方法を使用できます。プロセスは比較的簡単です。まず、小売業者は共有を作成する必要があります。共有は、デー

タコンシューマーに権限を付与してデータへのアクセスを許可するデータベースオブジェクトです。

同じデータを複数のデータコンシューマーと共有する

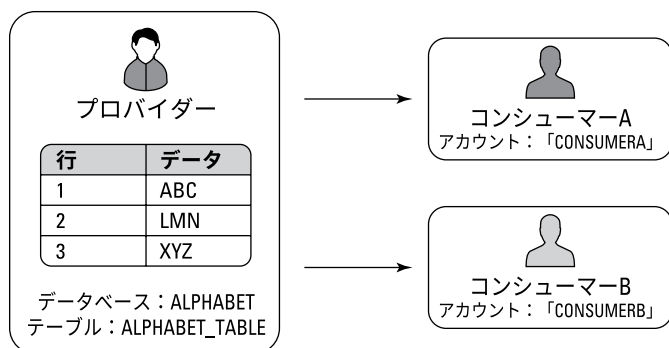


図 8-1: 権限を使って同じデータを複数のコンシューマーと共有する。

このプロセスを完了するため、小売業者は共有するすべての関連データベースオブジェクトに対する権限を追加してから、50件のフランチャイズアカウントを共有に追加します。この作業は、スケジュールされたスクリプトや、Python または SQL を使用したワークフローまたはプロビジョニングシステムによって自動化できます。

データのさまざまなサブセットを複数のコンシューマーと共有する

このシナリオは先ほどのシナリオよりも高度です。データのさまざまなサブセットを複数のコンシューマーと共有するからです。たとえば、ある自動車メーカーが、生産計画に関するデータを複数のディーラーと共有するとします。ディーラーは1つの州だけでビジネスを行っているとこもあれば、複数の州にまたがってビジネスを展開しているところもあります。この方法を使用すると、メーカーはディーラーの所在地にかかわらず、各ディーラーと関連データを共有できます。

たとえば、図 8-2 では、1つのテーブルのデータが2つのグループに分割されています。グループ1には行1が含まれます。これはコンシューマーAにのみ表示されます。グループ2には行2と行3が含まれます。これはすべてのコンシューマーに表示されます。

同じデータを複数のデータコンシューマーと共有する

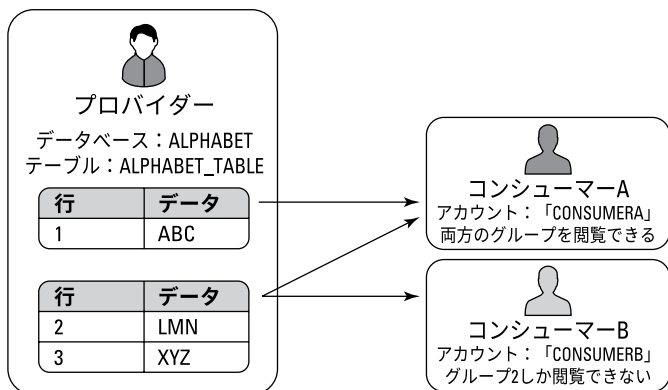


図 8-2: データのさまざまなサブセットを複数のコンシューマーと共有する方法。

この種の共有を簡単に行うには、いくつかの追加の手順が必要になります。まず、データプロバイダーが、ディーラーのアカウント名と各ディーラーがアクセスする必要があるオブジェクトのセットを対応付けるエンタイトルメントテーブルを作成する必要があります。次に、メーカーが、各ディーラーがクエリを実行する際にアクセスできるデータを制御するセキュアビューを作成する必要があります。

セキュアビューを使用して定義済みのデータスライスを提供する

第 4 章で説明したように、テーブルの一部にセキュリティや機密ポリシーが適用されている場合、テーブル全体を共有すると、機密データが公開される可能性があります。セキュアビューを使用すると、データへのアクセスを制御し、セキュリティ侵害の可能性を回避できます。

セキュアビューは、互いに外部の組織間でデータを共有する際にテーブル、列、行、さらにはセルレベルでセキュリティを適用するための効果的な手段です。

各データプロバイダーは、自身のデータのセキュアビューを作成し、そのビューへのアクセスを SaaS クラウドデータプラットフォームの他のユーザーと共有できます（これらのユーザーは別の組織に属していても構いません）。この方法は、複数の最終顧客に直接データベースへのアクセスを許可することを望んでいる組織にとっても有用です。セキュア

ビューでは、これらの顧客には各テーブルの特定のデータ行のみが表示され、他の顧客に関するデータが含まれる行は表示されません。

では、どのような場合にセキュアビューを使えばよいのでしょうか？以下にいくつかのガイドラインを示します。

- » セキュリティを侵害することなくデータコンシューマーにデータへのアクセスを許可する場合
- » PII/PHI が含まれている場合
- » 複数のコンシューマーが同じデータベースへのアクセスを必要としているときにデータアクセスを制限する場合

データプライバシーを維持しながら質問に答える

セキュアビューは強力な機能と柔軟性を兼ね備えていますが、ユーザーが `SELECT*` クエリを実行し、表示されるデータをすべて閲覧したりエクスポートしたりすることを防ぐ機能はありません。この種のアクセスを軽減するには、セキュアな共有ユーザー定義関数 (UDF) を作成します。これは、生データに対して安全に動作する小さな SQL または JavaScript コードです。セキュアな共有 UDF を使用すると、データの所有者は自身のデータに対して行うことができる質問や分析のタイプを制限できます。ユーザーに生データを直接表示する権限やエクスポートする権限を与えることなく、そのユーザーが詳細データに関する特定の質問を行えるようにすることもできます。

ある小売業者の例を考えてみましょう。その小売業者は、自社のサプライヤーがそれぞれの商品と一緒によく販売される商品を見られるようにしたいと考えています。これはマーケットバスケット分析と呼ばれる一般的な手法です。ユーザーに生データが表示されないようにしながらこのタイプの分析を実行できるようにするには、SQL ステートメントをセキュアな UDF でラップし、入力パラメーターでマーケットバスケット分析の対象として選択されている商品番号を指定します。

セキュアな共有 UDF を使用すると、データ所有者は詳細データに対して実行できる関数をデータコンシューマーに提供しながら、それ以外の関係者が生データを閲覧またはエクスポートできないようにすることができます。

データプロバイダーは、複数のコンシューマーにセキュアな共有 UDF を使用する権利を付与できます（それらのユーザーが別のアカウントや別の組織に属していても構いません）。セキュアな UDF は、データのセキュリティとプライバシーを保証するための優れた手段になっており、データプロバイダーとコンシューマーに以下のメリットをもたらします。

- » コンシューマーは、ソースデータや基盤ロジックを表示せずにクエリを実行できる
- » コンシューマーが尋ねられる質問、つまり実行できるクエリのタイプを制限できる
- » プロバイダーは、機密情報や PII が公開されることを心配せずにコンシューマーに価値あるデータを提供できる。

セキュアな結合を使用して企業間のデータを結び付ける

組織外の関係者とデータを共有する場合、データプロバイダーは共通の顧客を持つ複数の組織間でデータを共有していることに気付くことがあります。このような状況は、医療費支払者と医療機関がそれぞれ独自の患者データを管理している場合によく起こります。データを共有するときに、共通の患者のみに関するデータが各医療企業に表示されるようにするにはどうすればよいのでしょうか。HIPAA のガイドラインに従って PII を公開せずにデータへの双方向アクセスを提供するにはどうすればよいのでしょうか。患者の氏名や PII または PHI を実際に共有または転送することなく、患者に関するデータを交換することはできるのでしょうか。

セキュアな結合を使用すると、双方のデータセットにある記録のみを共同で分析し、重複しない患者に関するデータが各データコンシューマーに表示されないようにすることができます。セキュアな結合の詳細については、[Secure Joins: How to Join Data Between Companies While Respecting PII](https://www.snowflake.com/blog/secure-joins-how-to-join-data-between-companies-while-respecting-pii)（セキュアな結合：PII に配慮しながら企業間のデータを結合する方法）」（www.snowflake.com/blog/secure-joins-how-to-join-data-between-companies-while-respecting-pii）をご覧ください。



ポイント

包括的なデータシェアリングサービスは、[セキュアビュー](#)、[セキュアな共有 UDF](#)、[セキュアな結合](#)を使用してセキュリティを確保し、政府の規制を遵守しています。

- » 機会を評価し、組織の役割を定義する
- » データシェアリング機能を確認する
- » 万全の準備をする
- » 概念実証を実施し実行を成功させる

第9章

最新のデータシェアリングでビジネスを前進させるための6つのステップ

以上で、最新のデータシェアリングの大きな可能性と従来のデータシェアリングの課題が理解できました。次に、最新のデータシェアリングがもたらすであろう影響とメリットについて考えてみましょう。本章では、組織が最新のデータシェアリングを導入してビジネスを前進させるのに役立つ6つの重要なステップについて説明します。

1. **データシェアリングの障壁とチャンス进行明らかにする。** このステップの目標は、現在および近い将来における組織内のデータシェアリング要件の概要をつかむことです。データを共有するには、現在のデータフローや作業プロセスをしっかりと把握しておく必要があります。

この情報を手に入れたら、最大の価値を生み出す可能性を持つデータを特定することに集中してください。最終的な目的は、組織の生産性を低下させ、IT、データウェアハウス、ビジネスアナリティクスの各部門のリソースを浪費している煩雑なデータシェアリングおよびデータ転送プロセスを明らかにすることです。このような障壁があると、データからビジネス上の価値

を引き出すのに時間がかかります。障壁を特定し、データシェアリングのビジネスプランを簡単かつ迅速に実行する準備を整えてください。以下の点を見ていく必要があります。

- **データ**：どのようなタイプのデータをデータウェアハウスに保存する必要がありますか。新しいデータはどれくらいペースで作成されますか。データをウェアハウスに移動する頻度はどれくらいですか。現在または将来、データプラットフォームを介してデータエクステンジに容易にアクセスできるようにしたいですか。
- **データフロー**：組織のさまざまな部門内、および外部の組織との間ですでに行っているデータシェアリングを特定します。これらのデータフローを詳しく調査し、データを生成しているのは誰か、データを利用しているのは誰か、それはどのような方法で行われているかを把握します。
- **作業プロセス**：データシェアリングまたはデータ転送の作業プロセスを現在管理しているのはどのチームですか。使用されているツールを特定し、それらがETL（抽出、変換、ロード）ツールであるか、データレプリケーションツールであるかを明らかにします。この情報とデータフローの詳細を組み合わせると、データシェアリングに関連する現在のコストを特定できます。
- **市場投入までの時間およびインサイト獲得までの時間に關する将来的な目標**：現在のツールと利用可能な専門知識を考慮せずに、現在および近い将来におけるデータシェアリングのニーズを把握します。

2. ユースケースごとに役割を明確化する。現在および将来の各データシェアリングのシナリオを特定し、事業部門または組織がデータプロバイダーになるか、データコンシューマーになるか、またはその両方になるかを明確にします。

- データシェアリングに関わる内外の組織を特定します。どのデータシェアリング関係にも少なくとも2種類の関係者（データプロバイダーとデータコンシューマー）が存在します。
- データプロバイダーとデータコンシューマーを特定してデータシェアリングに参加させ、全員が複数の地域とクラウドインフラストラクチャプロバイダーにまたがって最新のデータシェアリングの機能にアクセスできることを確認します。

3. データプラットフォームソリューションが最新のデータシェアリングを簡単かつコスト効率よく実現できることを確認する。最新のデータシェアリングには以下の性能があります。

- データ移動が不要：最新型データシェアリングでは、データの移動、ETL、ファイル転送を行うことなくデータを共有できます。データ移動は、データの共有に必要な煩雑な作業の大部分を占めます。データ移動を排除すれば、制限の少ないデータシェアリングが可能になります。
- リアルタイム更新：データプロバイダーがデータを更新するとき、変更がコミットされるとすぐにすべてのコンシューマーに変更内容が表示される必要があります。これはデータプロバイダーまたはコンシューマーの追加作業なしに達成されなければなりません。データコンシューマーは、このようなデータの価値を高める高水準の完全性を求めます。
- 個別のセキュアビューとセキュア関数：最新のデータシェアリングの役割は、データウェアハウス全体へのアクセスをすべてのデータコンシューマーに提供することだけではありません。機密データを保護し、一般データ保護規則（GDPR）やカリフォルニア州消費者プライバシー法（CCPA）などの厳格な業界規制を遵守しながら、きめ細かい制御によってデータコンシューマーが求めるデータのビューを容易に提供します。最新のデータシェアリングでは、1対1、1対多、多対多のデータシェアリング関係を確立できます。ツールを選択する際には、このような柔軟性を備えたものをお選びください。
- 独自のプライベートデータエクステンションを作成可能：データの公開、共有、アクセスを行うことができるユーザを制御する独自のプライベートデータエクステンションを簡単に作成できることが重要です。

4. 概念実証（PoC）を実施する。データシェアリングのオプションの調査、デモの確認、質疑応答、ベンダーチームとのミーティングを終えたら、できる限り早く PoC を実施する必要があります。PoC は、ソリューションのテストを実施し、そのソリューションが自社のニーズをどの程度満たせるかを確認するプロセスです。

さらに、現在データを使用して行っていること以外に何ができるかを検討してください。最新型データシェアリングを実現する最新のクラウド構築型データプラットフォームを使用すれば、そのシステムからどのような追加のビジネス価値が得られるのでしょうか。データの一部を収益化したい場合、それを達成

するにはどうすればよいでしょうか。PoCを策定する際には、現在および将来の要件と成功基準をすべて列挙してください。

たとえば、現在のデータシェアリング方法に関する主な不満がクエリの実行に時間がかかりすぎることだとしても、その問題にだけ目を向けないでください。PoCでは、新しいソリューションへのデータの移行のしやすさ、新しい構造化・半構造化データのロード、クエリの実行、複数のワークロードの処理など、重要性の高いすべてまたはほとんどの要件に関する仮定を検証する必要があります。

5. **アウトリーチ計画を策定する。** データコンシューマーを関与させるために必要な手順をまとめます。最新のデータシェアリングでは、1対1、1対多、多対多のデータシェアリング関係を容易に確立できるため、データの価値とデータコンシューマーが期待できるメリットを伝達する必要があります。
6. **実行一市場投入までの時間と価値実現までの時間を短縮できることを関係者に実証する。** PoCが完了したら、最新のデータシェアリングのメリットを関係者に実証します。データプロバイダーまたはデータコンシューマーとして、組織の時間と関連コストをどれだけ節約できるかを見積もります。組織にもたらされる生産性の向上を実証し、必要な場合は、データの収益化によって得られる可能性がある収益を予測します。最新のデータシェアリングと最新のクラウド構築型データプラットフォームがもたらすROIの可能性に関する全体像を把握できるようになります。そうすれば、データシェアリング機能がレベルアップし、組織にとっての新たなチャンスがもたらされるでしょう。

最新型データシェアリングで、新たなビジネスチャンスを開き深いインサイトを得る

事業部門間とビジネスエコシステム内でライブの管理されたセキュアなデータを共有している組織は、顧客により良いサービスを提供し、新しい市場機会をより迅速に見出しています。しかし、従来のデータ共有手法では、煩雑でコストがかかり、リスクの高いプロセスのせいで、限られた成果しか得ることができませんでした。さらに、最新のデータ エクスチェンジを介してデータを安全に収益化することもできません。本書では、最新のクラウド構築型データプラットフォームを介してデータを簡単かつ安全に共有する方法を明らかにします。

本書の内容...

- データシェアリングが100年前に登場した理由
- 初期の手法で可能なことが限定される理由
- クラウド構築型アーキテクチャがライブでセキュアに管理されたデータシェアリングを実現する方法
- データを統制し、収益化する方法
- データシェアリングの実際の活用事例
- ライブデータの共有を開始する手順

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！



Lawrence C. Miller氏は、25年以上にわたって情報テクノロジーに携わってきました。彼はこれまでに70冊以上のFor Dummiesシリーズの書籍を執筆しています。**David Baum**氏は、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。

ISBN: 978-1-119-84433-4

再版禁止

for
dummies[®]
A Wiley Brand

 電子書籍
も販売中



WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.