



データガバナンスの成功に必要な**5大構成要素**

Snowflakeのクラウドデータプラットフォームは、有効なデータガバナンスプログラムを効果的に支える基盤となります。



TABLE OF CONTENTS

- 2 はじめに
- 3 構成要素#1: データアーキテクチャ
- 4 構成要素#2: データ品質
- 5 構成要素#3: データ管理
- 6 構成要素#4: データセキュリティ
- 7 構成要素#5: データコンプライアンス
- 8 結論
- 9 Snowflakeについて

はじめに

世界中の組織のリーダーがデータガバナンスの重要性を認めています。データガバナンスプログラムは、依然として計画段階か導入途中で足踏み状態にあります。

Dataversityが2020年に発行したレポート¹をみると、プログラムは完備していると答えた企業の割合は、わずかに12%、プログラムを導入中であると答えた企業は38%、着手したばかりと答えた企業が31%でした。

歴史を振り返ると、データガバナンスは大部分が規制コンプライアンスの指示した結果であり、GDPRなどの規制が組織に押し付けたものでした。しかし最近では、意思決定やアナリティクスもデータガバナンスを推進する要因となっています。これに応じてデータガバナンスプログラムの要件も変化しています。これらのプログラムは、企業がガイドラインや指令に従う際に役立つだけでなく、データドリブン型の企業に変わる力にも、ならなければなりません。

Snowflakeのクラウドデータプラットフォームを利用すると、有効なデータガバナンスプログラムをサポートする適切な基盤が得られます。Snowflakeを使用すると、企業でのデータサイロの解消に役立ちます。またコンプライアンスを達成し、安全で整理されたデータに基づいて、より効果的に意思決定を下せるようになります。たとえば3つの主要なデータクラウド、伸縮性のある

ストレージ、コンピュート機能での利用可能性、データ暗号化、アクセス制御、トラッキング機能、外部データ管理ツールとの統合などが含まれます。ここでは、成功するデータガバナンスプログラムを構築する際に留意すべき5大構成要素と、Snowflakeの機能がそれにどのように役立つかをご紹介します。



erwin, Inc. による「2020 State of Data Governance & Automation (2020年のデータガバナンスおよびオートメーションの状況)」
協力: DATAVERSITY Education, LLC, 2020年1月

¹ <https://bit.ly/31Yppll>

構成要素#1: データアーキテクチャ

最近のGartner社の調査²によると、ビジネス分野でサイロ化意識の克服は、データガバナンスやアナリティクスガバナンスで最も難しい側面の1つです。柔軟性に欠けるレガシーデータアーキテクチャは、組織全体での情報の共有と周知の障害となり、データの孤立化を助長します。

またレガシーアーキテクチャは、企業で一貫した情報整理が難しい理由にもなっています。情報が整理されておらず、サイロ化しているため、データシステムのトレース、データのカatalog化、詳細なセキュリティモデルの適用など、どの分野においてもデータガバナンスを敷くことができません。

Snowflakeは、サイロの削減、安全なデータ共有、柔軟性という3つの手段で、データガバナンスの成功を支える適切なプラットフォームとなります。Snowflakeの一元型メタデータレポジトリは、Snowflake内のすべてのデータに対してシングル・ソース・オブ・トゥールズ（信頼できる唯一の情報源）を提供します。これによってデータの検索、表示、トレースを簡単に実行できるようになるほか、手動、自動、AIのいずれの方法でもデータを簡単に整理できるようになります。データのソースと粒度

に透明性のあるプラットフォームを使用することで、データの発見、データのインベントリ、データシステムのトレースが容易になります。Snowflakeのデータ共有技術を利用すると、企業がデータを管理下に置いたまま、データの一元保管とトレースを複雑にしないで、安全にデータを共有することができます。顧客は、Snowflakeのいずれかの**セキュリティおよびガバナンスパートナー**を通じてデータのカatalog化を実行できるほか、任意のデータモデリング技術を使用できるため、カスタマイズされたガバナンスアプローチにつながります。



² <https://gtnr.it/2ZSs39p>

構成要素#2: データ品質

データガバナンスとは、企業に入ってくるデータの品質と、組織全体を通じたデータの使用を監視することです。データスチュワードは、データの破損、不正確さ、古さ、あるいは文脈に沿わない分析などを識別する必要があります。さらに、ルールやプロセスを簡単に設定できることも求められます。データを信頼できるということが、さまざまなソースからの情報に基づいて意思決定を行うデータドリブン型の組織の要です。Dataversityレポートに参加した企業の58%が、ソースデータの品質を理解することが、組織のデータのバリューチェーンで最も重大なボトルネックの1つだと回答しています。同レポートによると、「データ品質を最適化するには、ビジネス用語とデータ

アセットを自動化して照合し、カラムレベルまでのシステムを文書化することが重要」とのことです。

Snowflakeを利用すると各企業は、サイロを解消することによって自社のデータの有効性と正確さに対する信頼度を高めることができます。Snowflakeのストリームとタスクにより、パイプライン内で基盤となるデータへの変更を特定し、管理できるようになります。インバウンドのソースデータが変化すると、Snowflake内にパイプラインを構築して特定のパターンを発見することができます。また、そのような状態が特定された場合にはタスクがトリガーされて、発生した変化をトレースし、必要に応じて依存プロセスを脱却するロジックを推進できます。Snowflakeの外部機能を利用すると、アドレスの標準化や自然言語処理などの

サードパーティのサービスをパイプライン内で呼び出すことができます。そしてプラットフォーム内のデータ品質を高く保つのに役立ちます。最後に、堅牢なデータ品質が求められる場合には、Snowflakeのデータ統合パートナーの大規模なエコシステムを活用することで、データ品質とデータの信頼度をさらに高めることのできる機能が得られます。



構成要素#3: データ管理

データガバナンスにおいて企業は、いくつかの重要な質問に回答する必要があります。それは、自社にはどのようなデータが存在し、どこに保存されているのか、誰にアクセス権があり、それがどのように利用されているのか、といった質問です。

データへのアクセス、統合、保存、転送、および分析のためにデータの準備のための戦略や方法を整えるといったデータインベントリを実行するには、データ管理がカギです。Forresterリサーチ³には、「有効なデータガバナンスはデータ管理の成熟から生じる」と記載されています。しかし一方でIDC⁴は、データ管理の不備に苦慮している組織が全体の半数にも昇ると指摘しています。

Snowflakeは、集中型のデータディクショナリを通じて、企業によるデータのシステムとアクセスのトレースを支援します。これは、情報スキーマとも呼ばれるものであり、メタデータ自体を可視化することで、Snowflake内のオブジェクトに対するアクセス、変更、移動の状況を把握し、インサイトに昇華することができます。Snowflakeパートナーを通じ、企業は最適化されたコ

ネクターでデータのカタログ化やその他の管理機能を実行することができます。たとえばInformaticaのエンタープライズクラウドデータ管理プラットフォームでは、クラウドやオンプレミスのデータソースを特定してそれらに接続し、データのプロファイリング、データのカタログ化（システムを含む）、分析のためのデータの統合、ステージング、管理を実行することができます。SnowflakeとCollibraを統合することで、企業はCollibraによる管理されたデータカタログを利用してデータに関する完全なビジネスコンテキストを提供できるようになります。これにより、ITチームやビジネスチームは信頼できる管理されたデータを一元的にまとめ、コラボレーションでの利用、共有およびアクセスを実現することができます。

³ go.forrester.com/blogs/data-governance-for-2020-and-beyond
⁴ bit.ly/2ZfpwHD

構成要素#4: データセキュリティ

企業内外のデータソースの増加に伴い、データの侵害も増加しています。データガバナンスは、データセキュリティの向上に欠かせません。データ管理の成功の場合と同じく、データセキュリティには、データがどこから来てどこにあるのか、誰がアクセスしているのか、どのように使用されているのか、どのように削除するのかといったトレーサビリティが重要となります。データガバナンスでは、ルールや手順を定め、企業の機密情報や顧客データの漏洩を防ぐことで、悪意のある者の手にデータが渡らないようにしますが、レガシーなプラットフォームでは、情報がサイロ化され、アクセスやトレースが困難になります。そのようにサイロ化されたデータは、スプレッドシートなどにエクスポートされたり、他のサイロ化されたデータと組

み合わせるために複製されたりすることが多く、すべてのデータの所在を把握することがさらに困難になります。

Snowflakeのクラウドデータプラットフォームには、トレーサビリティとアカウントビリティを実現するいくつかの機能が備わっています。Snowflakeでは、役割ベースのアクセスコントロールを採用することで、オブジェクトへのアクセスを細かく制御し、誰がどのオブジェクトにアクセスできるか、そのオブジェクトに対してどのような操作を実行できるか、誰がアクセスコントロールポリシーの作成や変更を実行できるかを特定します。ダイナミックデータマスキングでは、セキュリティ管理者がカラムレベルでデータの可視性を制限するポリシーを作成します。ここでは、未編集のデータ、部分的に編集されたデータ、完全に編集されたデータが存在することになります。さらに、外部のトークン化サー

ビスとの連携も含まれるため、セキュリティの層を追加することができます。安全なビュー機能と安全なユーザー定義関数を利用すると、管理者はエンドユーザーが非表示のデータや情報をマスキングするロジックへのアクセスを禁じることもできます。さらに、Snowflakeはエンドツーエンドの暗号化(E2EE)を提供しており、データを読み取れるのは認証されたユーザーとランタイムコンポーネントに限定されます。Snowflakeのクラウドデータプラットフォームやその基盤となるクラウドベンダーを含むサードパーティは、暗号化されていない保存データに接触しません。クラウドプラットフォームのセキュリティが侵害された場合でも、E2EEが保存されたデータを常に暗号化することで、該当データを保護できます。



構成要素#5: データのコンプライアンス

企業は多くの場合、GDPR、HIPAA、PCI DSS、米国のサーベンス・オクスリー法 (SOX法) といった規制ポリシーへの準拠が必要となった段階でデータガバナンスのことを考え始めます。Dataversity レポートでは、48%の企業がデータガバナンスの主な促進因子として規制コンプライアンスを挙げています。これらの規制では、組織がデータのソースから消去までをトレースし、アクセス歴のある人物を特定し、データがどこでどのように使用されているかを把握できる必要があります。データガバナンスは、データの所有権とアクセシビリティに関するルールと手順を定めるものです。定めがなければ、機密情報が盗用・改ざんされ、政府や規制当局からの罰金、訴訟や懲役刑などにさえ、つながるおそれがあります。

Snowflakeには、データの所有権やアクセス権のコントロールを設定できる機能があり、データガバナンスのルールや手順を導入することができます。その一例が、ダイナミックデータマスキングや安全なビューです。データプロバイダーは、Snowflakeの安全なデータ共有によって共有したデータの所有権を保持しており、いつでもアクセス権を取り消すことができます。データの

コントロールを維持することで、企業はより簡単にデータの適切かつ安全な廃棄を実践できるようになります。Snowflakeのデータ統合パートナーは、監査やトレースのためのデータシステムツールを提供することで、このようなコンプライアンスシナリオを支えます。

コンプライアンスとセキュリティは非常に重要であるため、Snowflakeはセキュリティおよびコンプライアンスレポートのポートフォリオを継続的に拡充しています。現行のリストには、ISO 27001、AICPA/SOC、SOC 1 Type 2およびSOC 2 Type 2、HIPAA、FedRAMP Authorized (Moderate)、PCI DSS、FISMA Moderate、NIST 800-171、FIPS 140-2、ARS 3.の認証が含まれています。



結論

データガバナンスは、組織として継続的に取り組むものであり、コンプライアンス、ビジネス、セキュリティに関する新たな課題に合わせて成長させていく必要があります。データインフラ、データ品質、データ管理、データセキュリティ、データコンプライアンスに重点を置くことで、これらの課題に対応できるようデータガバナンスプログラムを拡張し、適応させていくことができます。Snowflakeのクラウドデータプラットフォームは、企業の規制コンプライアンスの取り組み、データドリブン型への移行、あるいはその両方において、データガバナンスを成功させるための強力な基盤を提供します。





SNOWFLAKEについて

Fortune 500企業180社を含む数千の組織が、サイロ化されたデータの統合、データの発見と安全な共有、多様な分析ワークロードの実行のために、Snowflakeのデータクラウドを利用しています。データやユーザーがどこに存在するかに関係なく、データクラウドは複数のクラウドと地域にまたがり単一のデータ体験を提供します。Snowflakeのプラットフォームは、データクラウドの動力源となり、アクセスを提供するためのエンジンです。データウェアハウジング、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、およびデータ共有のためのソリューションを生み出します。すでに多くのお客様、パートナー、データプロバイダーが、データクラウドを使用して新たなビジネスを切り開いています。貴社もその仲間入りを楽しみませんか。詳しくはwww.snowflake.comをご覧ください。

