



SNOWFLAKEによる データレイクの 大幅な強化



データレイク：約束

10年以上前、企業が収集して分析する必要のある多種多様なデータ、増え続けるデータ量に、レガシーデータウェアハウスでは対応できなくなりました。このようなデータには、ERPやCRMシステムなどの従来のデータソースから、サードパーティのデータソース、ウェブログ、IoTデバイス、ソーシャルメディアネットワークなどの最新のデータソースから得られるすべてが含まれています。初期の頃のビッグデータ向けのソリューションはデータレイクだったと考えられます。ここでは、組織が構造化データと非構造化データをすべて1か所に集めて保存し、分析することができます。オンプレミスのソリューションから始まりクラウドへと移行したデータレイクに期待されたのは、コストと複雑さを最小限に抑えて問題を解決することでした。

データレイクの当初の見通し：

- ビッグデータの多様性、速度、ボリュームを利用する
- 多くのデータタイプをそれぞれのネイティブフォーマットで保存し、ビジネスコミュニティが報告やデータサイエンス目的で利用できるようにする
- 組織が、複数のデータソースから断続的に受け取るペタバイト単位の情報を検索し、精査し、分析できるようにする

このような見通しは現在に至るまで達成されることはなく、データレイクのROIは低いのが一般的です。

組織が価値を実現できていない理由

多くのデータレイクは仮想的なデータのゴミ捨て場と化しており、有益な分析結果が得られない可能性があります。多くのデータレイクが抱えている課題を図1にまとめました。

- クエリ性能が限定的で、サイロ化した未活用のデータ
- 複雑なデータパイプラインと性能の低下
- ガバナンスとセキュリティが不足しているデータ

アクセスしにくいデータと限られたクエリ性能

あらゆる形式の未加工データをデータレイクに取り込むことは非常に簡単ですが、そのデータからインサイトを得ることは、非常に困難です。実際にインサイトを得ることができるのは一部の技術的な専門家だけであり、最新のアナリティクスを必要とする企業の意思決定者の大半がそのメリットを享受できていません。データレイクを十分に活用するためには、データへのアクセスを民主化する必要があります。ビジネスマネージャー、財務アナリスト、経営陣、データサイエンティスト、データエンジニアリングの専門家など、あらゆるスキルレベルの人々が、必要なアナリティクスのためにシステムに簡単かつ適切にアクセスできる必要があります。

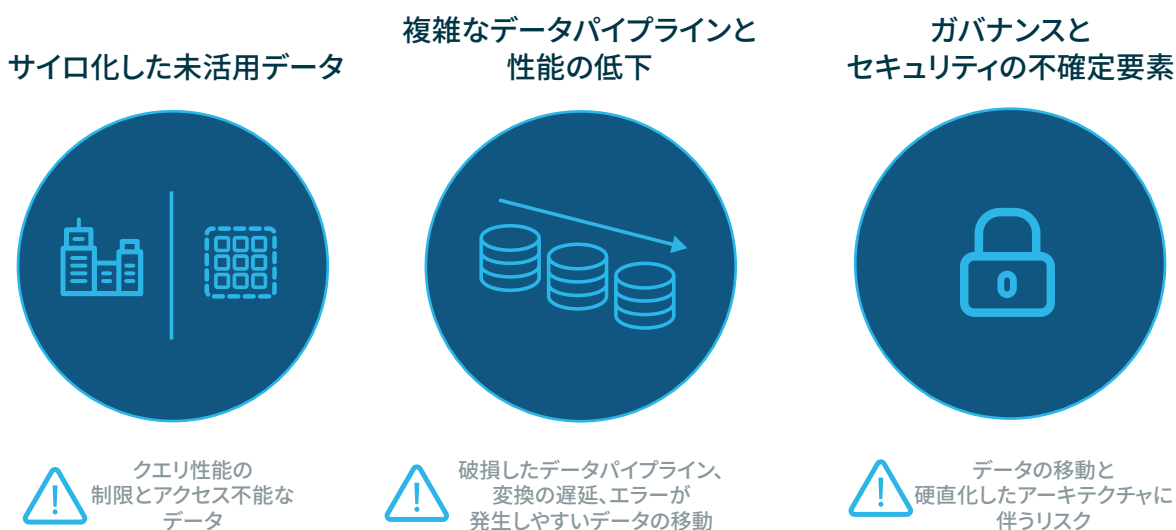


図1：データレイクに付随する課題

残念ながら、データレイクはともすれば、機構や構造のない無限に続くデータのリポジトリと化し、結果的にデータレイクが十分に活用されず、データの価値が十分に引き出されないままとなりがちです。

さらに同時実行性や拡張性の問題により、クエリ性能は多くの場合、非常に制限されます。そのためアナリストは、アナリティクス性能を制限されるだけでなく、データレイクではデータの大部分にアクセスできない事態に陥りがちです。

データパイプラインの破損と変換の遅延

データの大部分は、利用できるようになる前に変換が必要となります。しかし、データ変換には、さまざまなユーザーや使用事例のニーズを満たすために、多様なコーディングやツールのオプションの組み合わせが付いて回る場合が少なくありません。多くの場合、最終的なアーキテクチャが複雑になり、エラーの多いデータレイク外でのデータ移動を伴い、多くの技術のつぎはぎ細工で構成されることとなります。その結果、過酷なワークロードによってデータパイプラインが破損し、性能が低下しやすくなります。また、データエンジニアは、複雑なパイプラインの構築や、設定項目の多いインフラストラクチャの構成や管理にも時間を費やすこととなります。

ガバナンスとセキュリティが不足しているデータ

GDPRなどの規制やその他の地域のプライバシーに関する要件により、企業はデータ（その品質、使用事例、監査証跡、アクセシビリティなども含む）を正確に管理し、追跡し、理解することを余儀なくされています。しかし、その実践は困難を伴います。組織は、厳しいセキュリティとガバナンスのルールを適用した厳格なデータアーキテクチャを採用するか、より多くのユーザーが業務に必要なデータを活用できるよう厳格さを抑えるかという難しい選択を迫られています。適切なセキュリティとガバナンスを備えた柔軟なアーキテクチャを提供し、ユーザーが必要なデータを活用できるようにすることは、企業にとって非常に困難な課題となっています。

「2022年までに
ビジネス上の成果を
上げられる分析的
インサイトはわずか
20%に過ぎません。」

Andrew White
著名なVPアナリスト
Gartner
2019年1月

複雑なデータアーキテクチャ

従来のデータアーキテクチャは、最新のアナリティクスや効率的なデータパイプラインの要件に合わせて構築されていません。さまざまな形のデータがさまざまなシステムに格納されていることが多く、これらすべてのシステムを統合する簡単な方法がありません。また、リソースの競合によりパイプラインの性能と信頼性が低下します。さらに、複雑なインフラストラクチャを管理しなければならず、データパイプラインの構築と維持には時間がかかります。

図2は、多くの企業で見られる典型的なデータアーキテクチャを示しています。未加工のソースから目的の場所へデータを提供するために、テクノロジーの網目を複雑につなぎ合わせた図です。統合、変換、集約、提供の各段階で、組織は複数のツールを展開しており、それぞれの接続ポイントが障害点となる可能性があります。この

アプローチでは、ガバナンス、セキュリティ、サイロの問題が頻繁に起こります。なぜならビジネスのさまざまな部分で多様なニーズを満たそうとすると、結果的に複数のデータコピーが生じるからです。

最終的に、組織はインフラストラクチャの管理に多くの時間を費やすことになり、実際にデータを扱う時間はほとんどありません。ITチームが抱える技術的な負担は膨大であり、ビジネス上の要求に迅速に対応することができません。アナリストや企業的意思決定者が利用できるデータは不完全で古いデータに限られ、結果として信頼性に欠ける分析結果が生じることになります。



図2: 従来のデータアーキテクチャは複雑でコストがかかり、制約されています

データレイクを手放さず SNOWFLAKEで大幅に強化しましょう

求められるクエリ性能、データパイプラインの性質、セキュリティやガバナンスの要件にかかわらず、Snowflakeのプラットフォームは、ワークロードやユーザーのニーズを満たすのに十分な柔軟性を備えています。すでにクラウドデータレイクに投資し、投資を維持する予定の組織は、Snowflakeを使って既存のクラウドデータレイクを強化・補完することで迅速にROIを達成できます。さらに、次のようなメリットもあります。

- Snowflakeをデータレイク上で直接強力なクエリエンジンとして使用することで、データに簡単にアクセスできるようになり、リソースの競合や同時実行の問題のない優れたクエリ性能が得られます
- Snowflakeを使用すると、統合され、かつ拡張可能なデータパイプラインで、高性能で信頼性の高いデータ変換を実行し、データを簡単にデータレイクに再度アップロードすることができます
- Snowflakeのプラットフォームに組み込まれているガバナンスとセキュリティにより、データを扱うコラボレーションを拡充できます

迅速なクエリレイヤーとしてSNOWFLAKEを利用

クラウドデータレイクから迅速にROIを得るための簡単な方法は、既存のデータレイク上の迅速なクエリレイヤーとしてSnowflakeを利用することです。Snowflakeを使用して、データが存在する場所（つまり、Amazon S3、Microsoft Azure Blob Storage、Google Cloud Storageなどのクラウドストレージに格納されている場所）でデータをクエリすると、マルチクラウド環境であっても、データレイクをシングル・ソース・オブ・トゥールズ（信頼できる唯一の情報源）として維持することができます。シングル・ソース・オブ・トゥールズを持つことで、複数のデータリポジトリを同期するという時間のかかるタスクが不要となり、意思決定者が簡単にデータにアクセスできるようになります。

図3と次ページのテーブルに示すように、Snowflakeはデータレイクのデータをクエリして意味を理解し、シングル・ソース・オブ・トゥールズとして維持し、データの可能性を最大限に引き出す複数の方法を提供します。

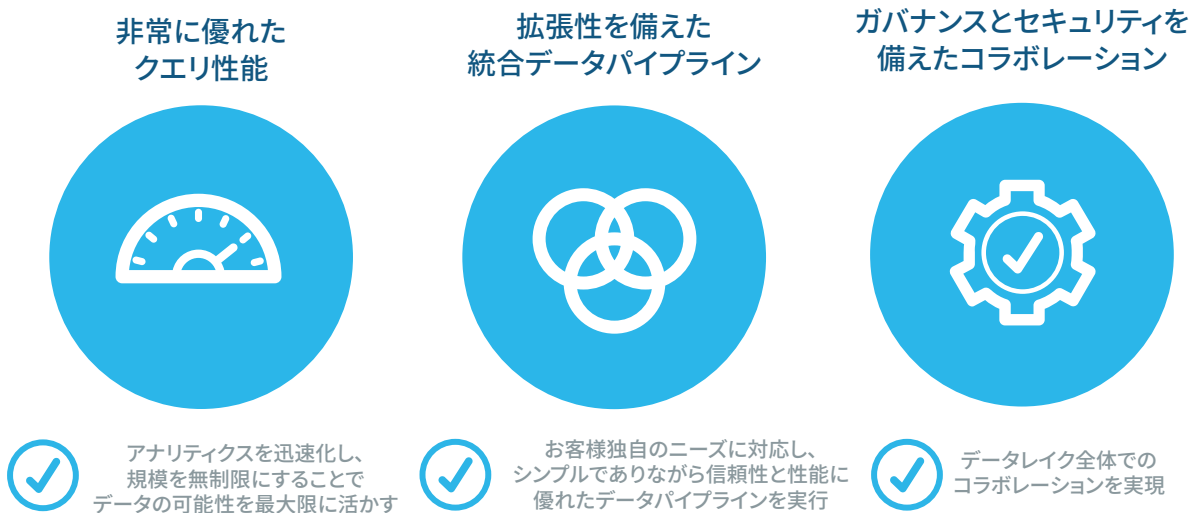


図3: Snowflakeにより既存のデータレイクを最大限に活用できるようになります

使用方法:	実行内容:
外部テーブル	データレイクで直接データをクエリ
お使いのクラウドの通知サービス	外部テーブルのメタデータを自動的に更新
外部テーブルのマテリアライズドビュー	頻繁に使用されるクエリの迅速化
パーティションの自動更新	新規ファイルの自動登録
Snowsight (Snowflake UIに埋め込み)	内蔵のチャートおよびダッシュボードでデータ探索を加速

外部テーブルがデータレイクを直接クエリ

外部テーブルとは、お使いの外部クラウドストレージにあるファイル上の読み取り専用テーブルです。外部テーブルを使用すると、Snowflakeにデータを移動することなく、データレイク内のファイルを直接クエリできます。外部テーブルには、ファイルパス、バージョン識別子、パーティション情報など、データファイルに関するファイルレベルのメタデータが保存されています。そのため、データレイク内のファイルに保存されたデータを、単一のデータベース内にあるかのようにクエリできるようになります。

外部テーブルは読み取り専用のため、DML操作は実行できませんが、クエリや結合には使用できます。外部テーブルに対してビューを作成することもできます。

外部テーブルにはデータファイルに関するファイルレベルのメタデータが保存されているため、データを含む実際のテーブルの場合と同じように、これらのファイルに保存されているデータをクエリすることができます。特定の場所に保存されているデータを読み取り専用で分析に利用しなければならない状況では、これが非常に好都合となります。データセット全体を実際のテーブルに読み込んでデータベースのスペースを圧迫しないで外部ステージでユーザーが変換、計算などを分析できます。

外部テーブルでは、たとえば以下を実行できます。

- Amazon S3、Microsoft Azure Blob Storage、Google Cloud Storageからデータを直接的にクエリする
- 自社のデータレイクをシングル・ソース・オブ・トゥールズとして維持し、データのコピーや転送の必要性を排除する
- データが格納されている外部ソースからデータを迅速に分析する

外部テーブルの自動更新で最新の状態をキープ

外部テーブルのメタデータに更新が必要になるたびに、お使いのクラウドの通知サービスからSnowflakeにその旨が通知され、Snowflakeが自動的に更新を行います。以下のような場面で、Snowflakeへの通知が行われます。

- テーブルのメタデータにパス内の新規ファイルが追加された場合
- パス内のファイルに変更があり、テーブルのメタデータ内で更新された場合
- テーブルのメタデータからパス内に存在しないファイルが削除された場合

次ページの図4に示すように、外部データは常に最新状態に保たれ、ユーザーがクエリや直接アクセスを実行できるようになっています。

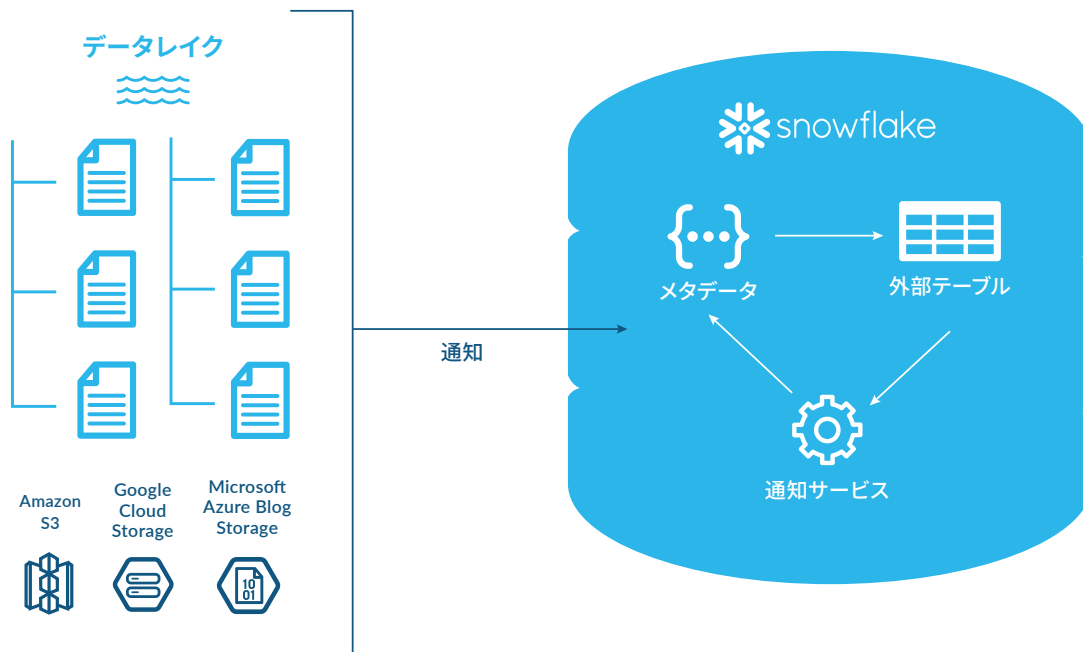


図4: 外部テーブルのパーティションが変更されたという通知をSnowflakeが受信すると、テーブルのメタデータが自動的に更新されます

マテリアライズドビューにより頻繁に使用されるクエリの迅速化が実現

データベースの外部に保存されているデータのクエリは、ネイティブなSnowflakeのテーブルのクエリよりも遅くなりがちですが、外部テーブルの上にマテリアライズドビューを作成することで、クエリ性能を大幅に高めることができます。マテリアライズドビューとは、(テーブルのように) Snowflake内に自身のデータを保存するデータベースオブジェクトです。別のオブジェクトのストレージを表示するだけのビューとは異なり、別途データを持ち、定期的に更新される別個のオブジェクトとなります。マテリアライズドビューは、大規模なデータセットに対するクエリ結果を迅速に提供し、ネットワークリソースの負荷を自動的に軽減します。

(外部テーブルの) マテリアライズドビューのクエリでは、元となる外部テーブルで同等のクエリを行うより高速性能が得られます。この性能の違いは頻繁に行われるクエリや、複雑性の高いクエリの場合に顕著となります。

マテリアライズドビューでは、同じサブクエリ結果を繰り返し利用するクエリの性能が改善する可能性があります。Snowflakeは、透明性を保ちながら自動的にビューを維持します。ベースのテーブルに変更が加えられ、バックグラウンドサービスによってマテリアライズドビューが更新されます。これは、マテリアライズドビューに相当するものをアプリケーションレベルごとに手動で維持するより効率的で、エラーの発生を抑えられる方法です。

ベースのテーブルでどれだけ多くのDML操作が実行されていたかにかかわらず、マテリアライズドビュー経由でアクセスされたデータは常に最新です。マテリアライズドビューが最新に更新される前にクエリが実行された場合、Snowflakeはマテリアライズドビューを更新するか、マテリアライズドビューの最新部分を利用してベースのテーブルから必要とされる最新データを読み出します。

マテリアライズドビューは、特に次のような場面で有益です。

- クエリに必要なスキンの行数を減らすための複雑で動的なフィルターや結合述語がある場合
- クエリ結果を満たすために必要なカラムの数が非常に限定的なクエリの場合
- クエリ結果に、次のようなインテリジェントな処理が必要な結果が含まれる場合
 - 半構造化データの分析
 - 計算に時間のかかる集約
- クエリの範囲が外部テーブルにある (つまり対象データセットが外部ステージのファイルに保存されている) ため、ネイティブのデータベーステーブルのクエリより速度性能が落ちている可能性がある場合
- ベーステーブルや外部テーブルに対するクエリに、集約や共通フィルター述語が含まれており、エンドユーザーの性能要件を満たしていない場合
- ビューのベーステーブルの回転が比較的小さい場合

以下のすべての項目が満たされる場合は、マテリアライズドビューを作成するとよいでしょう。

- 事前処理データ（集約、クラスタリング、外部テーブルからの取得など）により、エンドユーザーが実行するクエリの性能を大幅に改善される場合
- マテリアライズドビューのクエリや利用が頻繁な場合
- クエリ性能を高めることによるコスト削減（クエリのSLA要件を満たすために必要なウェアハウスの縮小）が、事前処理およびマテリアライズドビューの維持のコストを上回る場合

SNOWSIGHTによるデータ探索の加速

Snowflake UIには、データエンジニアのサポートを受けず自力でクエリ結果を確認したいユーザー向け設計された、SQLベースのデータ探索ツールが組み込まれています。Snowsightを利用すると、組織全体のコラボレーションが可能となります。クエリやダッシュボードを組織内の他のメンバーと共有しつつ、細かな許可を制御下に置くことができます。Snowsightでは、データからより迅速にインサイトを得るためにSnowflakeに組み込まれた充実した視覚化機能も利用できます。

Snowsightには次のようなメリットや機能が含まれます。

- 迅速かつ反応性に優れたクエリ：Snowsightは、高速のデスクトップ品質のエディターを提供します。
- スマートな自動入力：Snowsightを利用すると、クエリおよびSQLダイレクトに基づく文脈にそった提案が得られます。
- インタラクティブな結果：行やクエリの回答の多さにかかわらず、迅速にデータをプレビューできます。
- 視覚的に分かりやすいチャート：どのサイズのデバイスでも見やすいチャートが生成されます。
- 最新のダッシュボード：Snowsightでは、シンプルなドラッグ&ドロップインターフェースでダッシュボードを作成できます。
- ダイナミックなデータフィルター：パラメーターを使用して日付、顧客などを設定します。
- 共有可能なプライベートリンク：同僚に、コンテンツの閲覧、実行、編集が可能なリンク付きのクエリまたはダッシュボードを送信します。

すべてをまとめる

図5に示すように、Snowflakeのクエリエンジンは、既存のデータレイクを補完し、そこに保存されているデータへの迅速かつ簡単なアクセスを実現します。メタデータが変更されると自動的に更新される外部テーブルを介してデータレイクにクエリを実行したり、照会頻度の高いクエリをマテリアライズドビューの活用によってスピードアップしたり、SQLベースのSnowsightツールを使用してデータの意味を視覚的に理解したり、可視化したデータを同僚と共有したりすることができます。

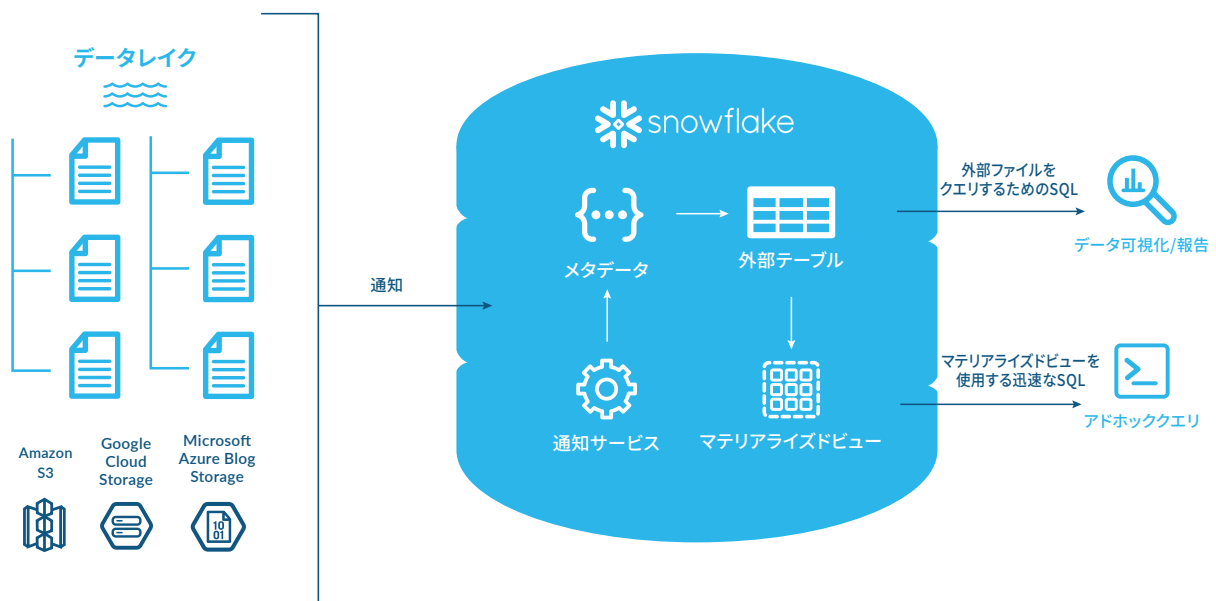


図5：強力なクエリエンジンとしてSnowflakeを使用することで、迅速なデータ分析が可能となり、既存のデータレイクへのアクセスが得られます

強力な変換エンジンとしてSNOWFLAKEを利用

クラウドオブジェクトストレージは、データレイクのデファクトの選択肢となりました。データレイクには大量のデータを保存できますが、そのデータを変換してインサイトを得るのは困難です。Snowflakeは、データレイクを補完するシンプルなアーキテクチャにより、ROIを迅速に改善し、データ変換に応えることができます。

アーキテクチャの設計方法によっては、外部テーブルを使用してクラウドデータレイクに直接アクセス

できます。あるいは、Snowflakeのサーバーレス取得サービスであるSnowpipeを利用して、データを自動的にSnowflakeに取得することもできます。その後、Snowflakeテーブルのストリームとタスクが変更データを取り込み、新規または変更されたデータを継続的に処理します。変換が完了したら、Snowflakeのデータレイクエクスポート機能を使い、選択した列で自動的にパーティショニングされたデータレイクに、データをアンロードすることができます。(図6を参照してください。)

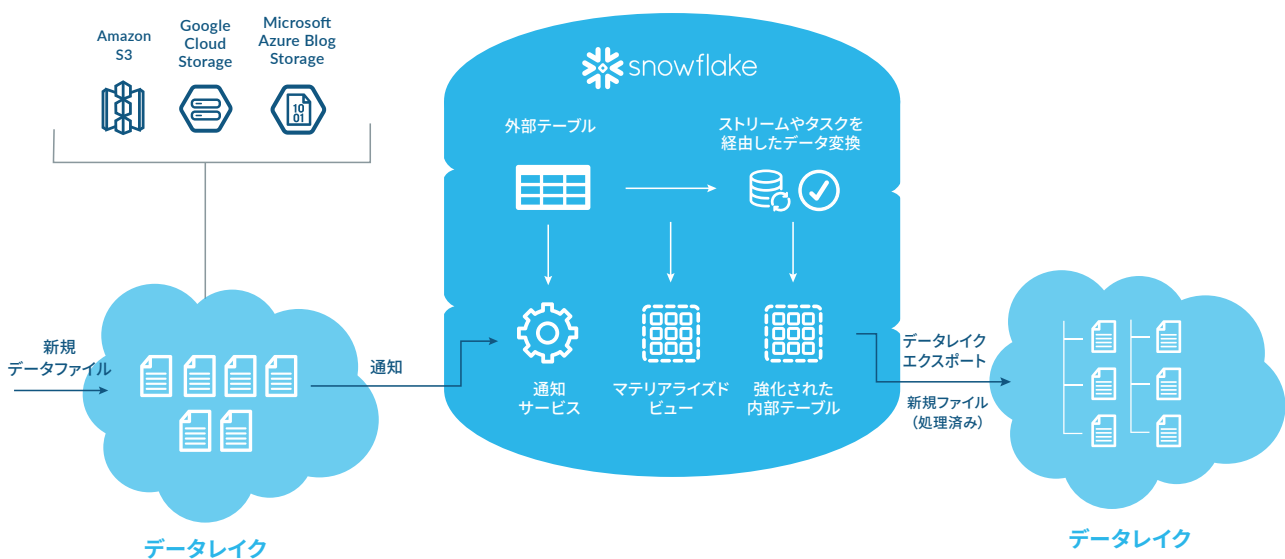


図6: Snowflakeを変換エンジンとして使用し、データレイクを大幅に強化します

継続的なデータパイプラインでデータフローを自動化

Snowpipeのサーバーレス取得サービスは、バックグラウンドサービスを使用して、クラウドストレージからデータを非同期的に読み込みます。データ取得にSnowpipeを利用するメリットを次ページの図7に詳しく示します。

CDC機能が変更内容を追跡

Snowpipeで継続的なデータパイプラインを構築することも、シンプルに外部テーブルを使ってクラウドストレージのデータを探し、Snowflakeのストリームとタスクを使ってデータ統合ジョブのスケジュールを設定し、変更データを取得することもできます。

- ストリームは、Snowflakeのオブジェクトタイプであり、挿入やDMLの変更など、テーブルの変更の差分を追跡するCDC機能を提供します。そのため変更後のデータを使い、アクションを起こすことができます。テーブルストリームでは、2つのトランザクション間でテーブルをクエリし、テーブルに対する行レベルの一連の変更を利用できます。
- タスクはSnowflakeのオブジェクトタイプであり、SQLステートメントを実行するための定期的なスケジュールを定義します。これにはストアプロシージャを呼び出すステートメントも含まれます。タスク同士をつなげて連続実行すると、より複雑な反復処理に対応することができます。

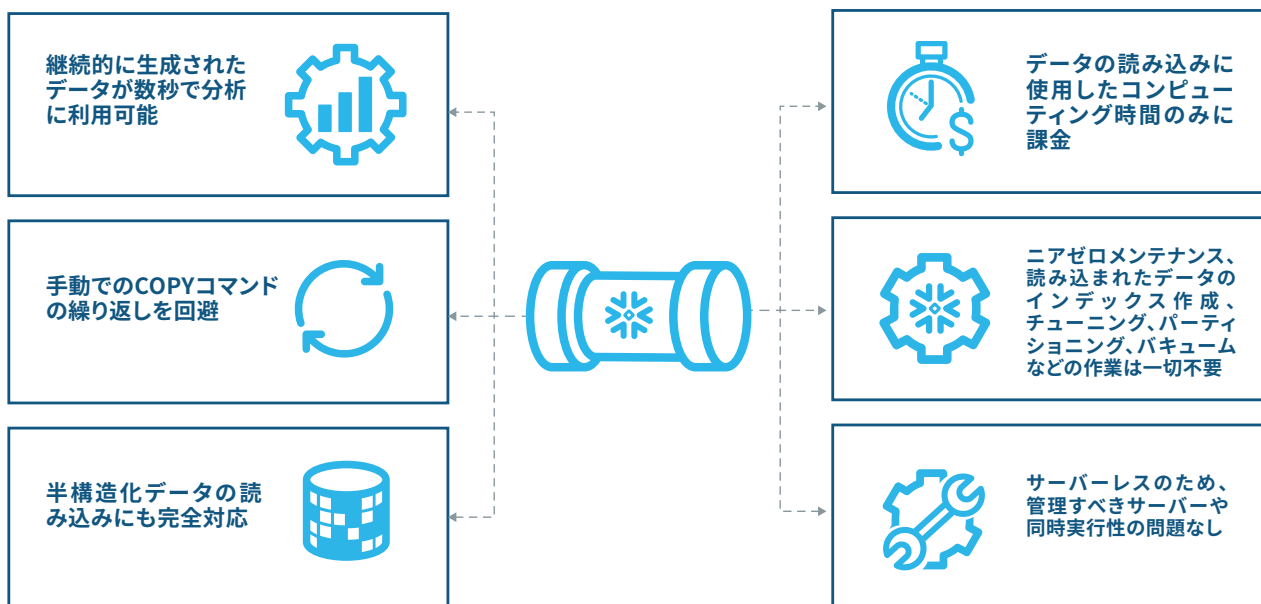


図7: SnowpipeはオブジェクトストアからSnowflakeへファイルを移動するプロセスを簡略化します

継続的なデータパイプラインでは、タスクはオプションでストリームを使用して新規または変更されたデータを継続的に処理する便利な方法を提供する場合があります。タスクは、ストリームにテーブルの変更データが含まれているかどうかを検証し、変更データを利用することも、変更データが存在しない場合は現在の実行をスキップすることもできます。

変換が終了すると、変換されたデータを、データレイクエクスポート機能を使用してデータレイクに再度アップロードすることができます。Snowflakeのストリームとタスクを利用すると、簡単にデータ統合ジョブのスケジュールを設定し、変更データを取得することができます。そのため、毎回すべてのデータを読み込む必要がなくなり、変更データのみを処理するだけでOKとなります。

ストアプロシージャによる頻繁なタスクの自動化

実行頻度が高く、複数のSQLステートメントを必要とするタスクは、プロシージャで自動化できます。プロシージャは、1回作成するとその後何度も実行することができます。

ストアプロシージャを作成するにはJavaScriptと、多くの場合はSQLを利用します。JavaScriptが制御構造（分岐とループ）を提供し、JavaScript APIの機能呼び出すことでSQLが実行されます。

ストアプロシージャで提供される内容は次のとおりです。

- 手続き型ロジック(分岐やループ)。ストレートSQLはサポートなし
- エラーハンドリング
- SQLステートメントを動的に作成し実行する機能
- プロシージャを実行する役割の権限ではなく、プロシージャを所有する役割の権限で実行されるコードを書く機能

拡張可能なデータパイプラインにより堅牢なデータ変換を実現

データ変換の多くは非常に複雑で、構築するのが困難です。たとえば、他の言語やフレームワークを使って構築する場合もあれば、サードパーティのコードや外部サービスを活用したりする場合があります。以前は、複雑なアーキテクチャを持つさまざまなデータ環境で、さまざまなサービスやシステムを管理する必要がありました。今では、Snowflakeではパイプラインを拡張することができるため、エクスターナルファンクションの定義や、サードパーティサービスを活用することができます。

ユーザーは、自身のリモートサービスの書き込みや呼び出しのほか、サードパーティ製のリモートサービスを呼び出すこともできます。これらのリモートサービスの書き込みには、AWS Lambdaなどのクラウドサーバーレスコンピューティングサービスを含め、あらゆるHTTPサーバースタックを使用できます。

エクスターナルファンクションの例をいくつか紹介します。

- ジオコーディングサービスを利用して住所を座標や政治的地域で増補する
- サードパーティサービスを利用してメッセージのセンチメント分析を実行する
- カスタム機械学習モデルを利用して顧客をスコアリングする
- カスタムロジックを利用してメールからメールアドレスを抽出する
- リモートサービスからライブの株価を取得する

マルチクラスターのコンピューティングアーキテクチャが伸縮性と拡張性を提供

Snowflakeのデータパイプラインは、バッチまたは継続のいずれかを選ぶことができ、処理はSnowflake自体の中で直接行われます。Snowflakeではマルチクラスターのコンピューティングアプローチを採用しているため、これらのパイプラインはその他のワークロードの性能に影響を与えることなく複雑な変換に対応できます。

その仕組みについて説明します。Snowflakeのマルチクラスターでは、共有データアーキテクチャ、コンピューティングおよびストレージリソースは物理的には別物ですが、論理的にはこれらが単一の統合データウェアハウスシステムを構成していることとなります。この独自のマルチクラスターアーキテクチャにより、ユーザーが必要なだけ個別のワークロードをサポートできます。各ワークロードに専用のコンピューティングエンジンがあり、必要に応じてスケールアップやスケールダウンができます。事前にリソースを割り当てておいたり、他のプロセスを中断したりする必要はありません。

この革新的なアーキテクチャにより、Snowflakeは複数の多種多様なワークロードに簡単に対応することができます。コンピューティングとストレージが分離されているため、ELTプロセスの実行や、BIレポートユーザー、データサイエンティスト、データマイナーのサポートのために、異なるサイズの仮想ウェアハウスをスピンアップすることができます。この場合のリソースの競合はまったくありません。

半構造化データをネイティブにサポートし変換アーキテクチャを簡略化

Parquet、ORC、JSON、XML、Avroといった半構造化データ形式は、厳密には型指定されていません。通常はノードが階層化されています。ノードごとに名前1個

と値1個があり、値の内訳は定数値1個、名前と値のペアの列1個、ネストされたオブジェクト1個のいずれかです。従来のデータベースでは、一般的に、異なる構造を扱うことができないため、開発者はETLツールに頼り、データを従来の行と列の構造にフラット化する複雑な変換を設計することになります。

Snowflakeは、VARIANTという半構造化データに特化したデータタイプを導入しました。ここで、半構造化データの取得プロセスの仕組みをご紹介します。仮想ウェアハウスは、JSONまたはParquetのドキュメント全体を、そのままの状態でも VARIANT列に書き込みます。ユーザーがSQLでその列をクエリすると、同じJSONまたはParquetのドキュメントが戻ります。

同時にグローバルサービス層がドキュメントのコンテンツに関する特定のメタデータ（ノード名、ヒエラルキー、配列構造）を収集し、そのデータをメタデータストアに書き込みます。Snowflakeは、ユーザーがSQL内で直接データを参照できるよう、内部でデータを列に整理します。この列データはマイクロパーティションファイルに書き込まれますが、テーブル定義には実際の列名は表示されません（JSONまたはParquetノードの列名は存在しないということになります）。VARIANT列のノードを参照するには、SQLの単純なドット記法を使用します。このプロセスは性能が高く、明示的な変換を記述する必要がありません。

統制されたセキュアなデータのために SNOWFLAKEを利用

データがどこにあっても、Snowflakeはグローバルなデータガバナンスとセキュリティを確保します。粒度の高いセキュリティ、ロールベースのアクセスコントロール（RBAC）、安全なデータシェアリングの提供が難しい従来のソリューションとは異なり、Snowflakeで以下のような堅牢なガバナンス機能がプラットフォームに搭載されています。

- RBAC
- エンタイトルメントテーブルによる行レベルのセキュリティとアクセスコントロール
- マスキングやトークン化のオプションによる外部データの列レベルでのセキュリティ
- セキュアなライブデータシェアリングとデータエクスチェンジ（外部テーブルによるクエリレイヤーとして）
- 半構造化データのセキュリティとガバナンス

この内蔵アプローチにより、セキュリティおよびガバナンスのチームは以下を実行できます。

- 精度の高いアクセシビリティのための詳細なアクセスコントロールを設定する
- 列レベルのマスキングや行レベルのフィルタリングなどの機能を活用し、きめ細かなセキュリティを実現する
- ユーザーが機密情報を漏らすことなくデータを扱う能力を最大限に引き出す
- 社内外の関係者との安全なライブデータシェアリングを行うことで、より効果的なアナリティクスとコラボレーションを推進する

あるいは…お使いのデータレイクの代わりにSNOWFLAKEを採用

お客様のデータのニーズに適合する場合は、Snowflakeをデータレイクやデータウェアハウスとして利用していただくこともできます。Snowflakeでは、未加工データステージを設定して、データストレージを3分の1～5分の1に圧縮することができます。それにより、前述の拡張可能性、性能、使いやすさなどのメリットに加えて、シンプルさも実現できます。1つのプラットフォームと1つのデータコピーだけで、さまざまなワークロードや使用事例に対応し、クラウド全体でエンドツーエンドのガバナンスを実現します。

既存のデータレイクインフラストラクチャをSnowflakeに置き換える際に、Snowflakeのデータクラウドを選択することもできます。データクラウドを採用すると、クラウドデータストレージやオンプレミスのデータセンターに隔離されたサイロ化したデータセットが解消され、組織は、簡単にすべてのデータを単一のコピーに統合し、それに接続できるようになります。データクラウドは何千もの組織が共有データやデータサービスをシームレスに共有し利用する形で、組織内のデータに接続し、組織間でも相互に接続するエコシステムです。

データレイク用のSnowflakeの詳細については、[こちら](#)をご覧ください。



SNOWFLAKEについて

Snowflakeは、Snowflakeのデータクラウドを用い、あらゆる組織が自らのデータを活用できるようにします。お客様には、データクラウドを利用してサイロ化されたデータを統合し、データを発見してセキュアに共有し、多様な分析ワークロードを実行していただけます。データやユーザーがどこに存在するかに関係なく、Snowflakeは複数のクラウドと地域にまたがり単一のデータ体験を提供します。多くの業界から何千ものお客様（2021年7月31日時点で、2021年Fortune 500社のうちの212社を含む）が、Snowflakeデータクラウドを自社のビジネスの向上のために活用しています。詳しくはwww.snowflake.comをご覧ください。

