株式会社NTTデータ ニューソン
2024 年 12 月

ビッグデータや AI 領域で活用が進む Snowflake の解説と導入事例の紹介

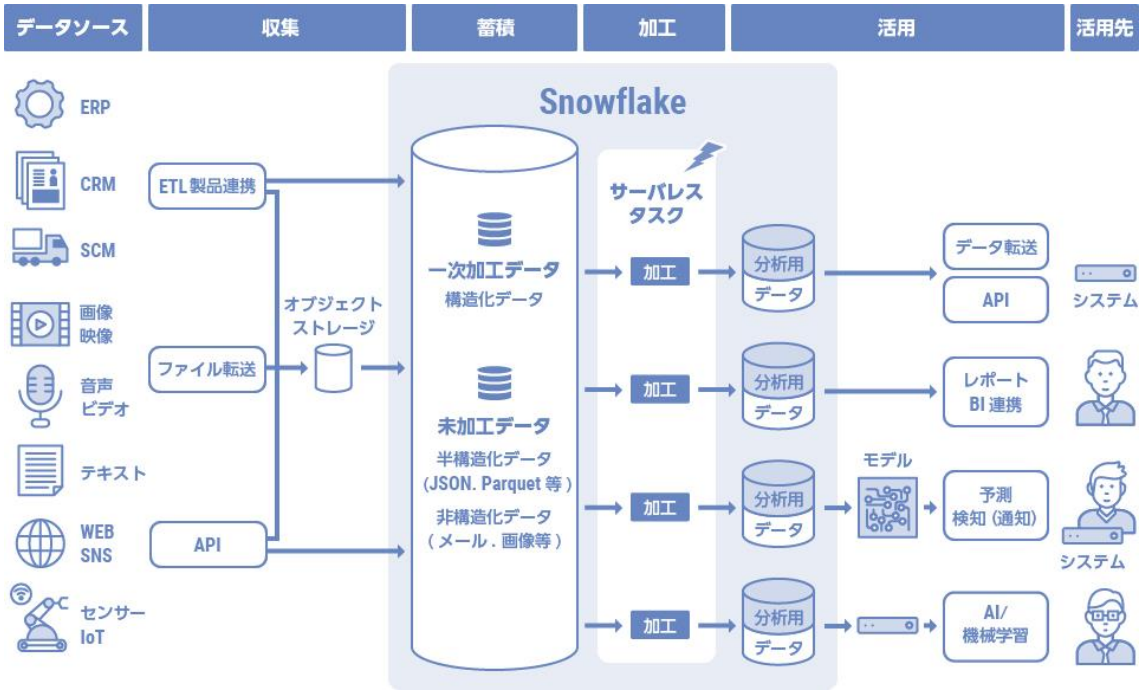
～ クラウドを活用した並列分散処理により、大量データ分析の時間とコストを大幅に削減 ～

各社がビッグデータの管理や活用に注力するなか、柔軟で高速なデータ処理が可能な「Snowflake」に注目が集まっています。本レターでは、Snowflake の特長と事例をご紹介いたします。

ビッグデータの分析や AI アプリの構築も可能な Snowflake の特長

Snowflake は、クラウドベースのデータプラットフォームです。大量のデータを一元管理し、高速に分析・処理できるのはもちろんのこと、Streamlit※1 を使ったアプリケーション開発も可能です。また、生成 AI サービスにも注力しており、「Coretex AI」から大規模言語モデル（LLM）へアクセスし、AI アプリケーションを構築できる機能を拡充するなど、活用の幅が広がっています。

※1 Python で Web アプリケーションを作成するためのフレームワークのこと



参考：Snowflake を使った活用図

パブリッククラウドの特性をいかした優れたアーキテクチャ

Snowflake のアーキテクチャは処理が実行される仮想ウェアハウスとデータを保存するストレージが分離しているため、複数の仮想ウェアハウスを同時に稼働でき、高負荷の状況でも複雑な処理を高速に実行できます。

また、データは小さく分けた状態で各パブリッククラウドの安価なストレージサービスに圧縮して格納されます。そして、どのように格納されたかについても整理されてサービス層にメタデータとして保持されます。この優れたアーキテクチャにより、ボトルネックが無く、様々な機能がリリースされても性能を維持することが可能となります。

運用・保守作業の負荷低減

ビッグデータ運用は、扱うデータの量が多く、データ保護にも莫大な費用を要すると思われていますが、Snowflake を活用することで運用や保守作業を低減できます。Snowflake には、特定の時点に戻れる機能や 1 つのゾーン障害に対応した保護機能が標準で提供されており、データ格納領域（マイクロパーテーション）の最適化も自動で運用できます。また、パブリッククラウドやリージョンの障害への対策として、レプリケーション機能が提供されており、比較的安価に BCP 対策することが可能です。

柔軟なデータ共有と管理

Snowflake アーキテクチャではデータをコピーせずに企業間での安全なデータ共有が可能です。また、多くの SaaS ベンダーが Snowflake データマーケットプレイスにデータセットを提供しているため、簡単な手続きでデータにアクセスすることができます。
多くのデータを取得することにより、多様化するビジネス環境への対応が可能になります。
また、データ格納を ELT 方式にし、生データを JSON 形式（VARIANT 型）のような半構造化データで一旦取り込むことにより、直近では利用しない項目値に関するデータ設計を先送りすることができます。
これにより、新しいデータ種を利用者に提供するまでの時間（本番移行までの時間）を、データ変換を行ってから格納する ETL 方式に比べて短くできます。
そして Snowflake では半構造化データ形式でデータを保持した場合でも、SQL のみならず Python や Spark を用いてクレンジング・加工が可能です。

フレキシブルな従量課金

従量課金制のため、目的にあわせて稼働を調整し、高いパフォーマンスを保ちながらコストを抑えることができます。
複雑な処理をする時には性能を高く、それ以外の時間帯は性能を低くしたり、稼働を止めたりするなど、利用状況に合わせたフレキシブルな課金が可能です。

Snowflake の特性をいかした環境構築で、工数やコストを削減

NTT データ ニューソンには Snowflake の技術認定資格である SnowPro Core を保有しているメンバーが複数在籍しており、Snowflake の特性をいかした活用方法を提案しています。データ分析基盤の構築や他システムから移行し、工数やコストの削減に繋げた事例等をご紹介します。

事例 1：故障診断システムの環境構築・データ分析・維持管理（製造業のお客様）

AWS フルマネージドサービスと Snowflake を使った診断環境の再構築をしました。
Snowflake には IoT ログが Snowpipe※2 により随時テーブルに取り込まれており、全体で 1 兆レコード超もの大量のデータを保持しています。
再構築前は、日々の分析バッチ処理をおこなうため、毎日数十億レコードを高価なサーバーに取り出し一晩中処理をおこなっていました。
再構築時には、データを抽出する際に Snowflake 側では Python を利用して前処理をおこない、AWS 側では Lambda で 1,000 並列に分散して診断処理をおこなうようにしたため、30 分から 1 時間程度で処理が可能になりました。また、高価なサーバーを使う必要がなくなり、費用も大幅に削減できました。
本プロジェクトの対応を始めた当初は、SQL や Python を活用した場合でも Snowflake で細かいデータ分析をおこなうのが難しい状況だったため、Snowflake と AWS フルマネージドサービスを組み合わせて処理をする作りとなっていました。しかし、Snowpark という機能により Snowflake 内部で Spark 処理ができるようになりました。今後は Snowflake でデータ分析を完結させ、さらなる費用低減への取り組みを進める予定です。

※2 Snowflake のテーブルに対して継続的にデータをロードする仕組み

事例 2：データ分析基盤および分析モデルの開発環境構築の支援（製造業のお客様）

ローカルに存在していた IoT ログを Snowflake にインポートする独自 Python ツールを開発しました。

多数の CSV、TSV が存在し、同じデータ種でも発生時期によりカラムの数が異なるという状態だったものの、ファイルにヘッダレコードが付与されていたため、インプットファイルを JSON 形式に変換する汎用的なツールを作成しました。それにより、ELT 方式で、短期間で Snowflake へ格納することを実現しました。また、汎用ツールの中でテーブル作成をする際、抽出条件で頻繁に利用するカラムは JSON 外でも独立したカラムとして保持するよう、設定ファイルでコントロールできるよう工夫しました。

分析モデルの開発環境構築においては、Snowflake と連携して特定期間の IoT ログを取得してクレンジング・前処理・可視化を行う Python の分析環境を構築するとともに、分析時間を更に短縮するために Snowflake 内で全て処理を行う Snowpark を使った Python の分析環境の構築も行っています。

事例 3：開発の生産性向上に向けたデータ加工・運用業務（通信業のお客様）

開発の生産性を向上させるために Snowflake と dbt を活用したデータ加工・運用業務の支援をしました。

お客様はサービス向上のため、利用情報などのデータ分析を迅速に行う必要があり、Snowflake を導入することで高速かつ安定したデータ分析基盤を実現していましたが、ビジネススピードが加速するなかで、データ加工におけるデータマートの開発生産性が課題となっていました。

そこで、データ変換に特化したツールである dbt を導入しました。dbt は、実行結果を自動でテストしたり、処理フローを視覚的に確認したり、仕様書を自動生成したりといった機能が充実しており、Snowflake 単体では補いきれない部分をカバーしてくれます。これにより、データ変換に関する作業を効率化し、開発のスピードアップを実現しました。

さらに、dbt.Labs 社が公開している dbt ベストプラクティスに則って 3 つのレイヤーに分けて設計・構築するなど統一ルールを設けることにより、開発者が増えると発生しがちな品質のばらつきや教育面の負担軽減も図っています。

今後も NTT データ ニューソンは、Snowflake の特性を最大限に活かす開発を支援してまいります。

【参考情報】

・NTT データ ニューソンの Snowflake サービスについて

https://www.newson.co.jp/services/DigiSol/Snowflake/

■ NTTデータ ニューソンについて

株式会社NTTデータ ニューソンは、情報システムの企画、設計、開発、保守をしています。1974 年に設立以来、システムインテグレータとしてソフトウェアとハードウェアの融合を行っております。2017 年にNTTデータグループの一員となり、2024 年に創立 50 周年を迎えました。

NTTデータ ニューソンは、NTTデータグループの各社と連携し、”情報技術で、新しい「しくみ」や「価値」を創造し、より豊かで調和のとれた社会の実現に貢献すること”を目指しています。

NTTデータ ニューソンに関する詳細な情報については、https://www.newson.co.jp/ をご覧ください。

＜ 本件に関するお問い合わせ先 ＞

株式会社NTTデータ ニューソン　営業推進室

E-mail：pr_newson@newson.co.jp