

Snowflakeでペタバイト級データ基盤を刷新

NTTドコモのIDAP、運用効率とデータROIを劇的に改善し 高速・効率的なデータ分析を進化

通信

NTT
docomo

企業 株式会社NTTドコモ R&Dイノベーション本部
所在地 東京都千代田区永田町2丁目11番11号
山王パークタワー

ご利用のSnowflakeワークロード



NTTドコモの「IDAP」は、通信エラーログをはじめとする通信事業に関するデータとスマートライフ事業のdポイントクラブ会員情報などを一元的に管理するビッグデータ分析システム。クラウドベースで運用されるペタバイト級データ基盤は、データ量と処理量の拡大に伴う管理工数の増大や、処理別コスト管理の煩雑さによるデータROI向上の困難さなどの課題に直面していた。同社は、データ基盤をSnowflakeに全面移行することで、保守・管理工数の大幅削減と分析単位のコスト把握を実現している。

このストーリーのハイライト:

- 巨大データ基盤の保守・管理工数を大幅に削減
- プロジェクト別のデータ分析コスト把握を実現
- 処理に要する時間のプレを大幅圧縮

課題:

スケールの困難さが生んだ巨大データ基盤の4つの課題

NTTドコモが2014年に運用を開始したビッグデータ分析システム「IDAP」(Integrated Data Analytics Platform)は、14PBのデータを保有し、1日200TB以上のデータ処理を行うペタバイト級データ基盤である。通信事業部門とdポイントクラブをはじめとするスマートライフ事業部門のデータを統合する巨大データ基盤は、すでに様々な成果につながっている。

2024年7月から基地局への適用が開始された、量子コンピューティング技術を用いた通信サービス最適化もその一つだ。モバイル通信では、基地局からユーザー端末にページング信号を発信し、端末からの応答を得ることで位置を把握する。スマートウォッチやIoT機器の増大に伴うページング信号増大は基地局の新たな負荷につながっている。それを受けて同社が目にしたのが、量子効果を活用した新しい計算手法である量子アニーリングだった。同社が開発した量子コンピューティング基盤は最大15%のページング信号削減を実現し、今後は通信だけでなく、金融や物流・流通など多様な分野の最適化問題に活かされることが期待されている。その開発に大

きな役割を果たしたのは、IDAPが保持する膨大なデータだったという。

さまざまな成果の一方で、複数のクラウドデータ基盤サービスを利用して構築されたIDAPは、利活用の進展に伴いさまざまな課題に直面していた。まず挙げられるのは、プロジェクトや施策ごとの分析コスト把握の困難さである。R&Dイノベーション本部 サービスイノベーション部 ビッグデータ基盤担当 担当部長の石井 啓之氏はこう説明する。

“「事業部門の活用が進んだことでまず問題になったのが、増加した分析コストをどう按分すべきかという問題でした。施策ごとにどれだけ使ったかが見えれば問題ないのですが、以前の環境ではそれが困難でした。これは結果として、データROI測定の困難さにもつながっていました。コストを抑えたいというよりも、このままでは非効率なクエリ発行が野放しになってしまうと感じました」

次が処理に応じた計算リソース変更の難しさだ。この問題はアドホック処理の困難さと共に、処理時間が予想できないというもう一つの問題にも直結していた。R&Dイノベーション本部 サービスイノベーション部 ビッグデータ基盤担当 Principal Data Engineerの松原 侑哉氏はそれをこう説明する。

「IDAPのユーザー数は現在3000人を越えますが、データ利活用の促進は結果として処理が集中する時間帯のレスポンスにつながっていました。例えば、通信事業部門では、通信エラーログを一定時間ごとに分析してネットワークの問題を抽出していますが、処理速度の遅れは高品質な通信サービス維持の困難さにもつながります」

解決策:

割り振ったリソースの自動スケーリングで費用を可視化

松原氏がSnowflakeの存在を知ったのは2020年頃のこと。ペタバイト級データ基盤の全面的な移行先として本格的な検討を開始したのは2023年

約100時間

毎月の保守・運用時間を大幅削減

最大1/5

旧環境と比較し、
高いデータ圧縮率を実現

2分

最大30分の処理時間のプレが
2分以内に収束

になってからのことだ。

「まず懸念したのは、SnowflakeがIDAPの処理量に本当に対応できるのかという点でした。負荷に応じて計算リソースをスケールする自動スケールがスムーズに動くことを確認し、移行可能と判断しました。また自動スケールは、課題の一つだった処理時間のブレ改善にも大きな役割を果たすことが期待できました」

コンピューとストレージを分離し、計算リソースのスケールがスムーズに行える特長はもう一つの効果にもつながった。プロジェクトごとに計算リソースを提供することで、事業部門別のコスト管理が可能になるからだ。

「プロジェクト別に計算リソースを用意することは以前の環境でも可能でしたが、分析をスムーズ行うには分析クエリ発行による負荷をあらかじめ把握する必要があるため、現実的な選択肢ではありませんでした。しかし、Snowflakeの自動スケールを利用することで、管理者の負荷を増やすことなく、プロジェクト別にリソースを提供することが可能になります」

結果：

可用性向上により、保守や障害対応工数を大幅削減

IDAPに向けた取り組みが本格的にスタートしたのは2023年6月。数十件のシステム間連携、数千件のワークロード/データ処理パイプライン、約5000テーブル/ビューのデータ移行を含む作業が急ピッチで行われ、翌年6月には旧データ基盤の提供を停止し、移行を完了している。松原氏はこう振り返る。

「技術的な難しさはまったくなく、データ移行に伴う加工と処理スクリプトの書き換えやシステム間連携の調整が必要になり、多くの人員を投入して対応する必要があったことが一番の苦労でした。既存環境から移行までの期間が限られていたこともあり、時間との闘いではありましたが、特にトラブルもなく移行を終えています」

IDAPは、アクセス元のセキュリティレベルに応じ、生に近いデータ(セキュリティレベルS)と統計化されたデータ(セキュリティレベルB)の二通りのデータを用意し、セキュリティレベルSのデータについては各拠点の専用端末からのみアクセスできるようにし、公開用/作業用データベースを分離するなど、万全のセキュリティ対策を図っている点も特徴の一つだ。Snowflakeへの移行にあたっては、同じ役割のデータベースRoleを同一名称で各データベースに格納し、オブジェクト命名規則を徹底するなど、管理業務の省力化に向けた様々な工夫も実行している。

SnowflakeのDynamic Data Masking / Row Access Policy機能による、アクセス権限に基づいたデータ閲覧制御もその一つだ。それにより、一つのテーブル/ビューで二つのセキュリティレベルに確実に対応することを実現している。

松原氏が導入効果としてまず挙げるのは、保守・運用・障害対応に要する時間の大幅な削減である。

「IDAPはデータ量、処理量共に膨大だったこともあり、これまで世界的にも稀なトラブルも含め、様々な障害に直面してきました。Snowflake移行後は、21名のスタッフが障害に対応する時間は月100時間削減できています。またリソース自動最適化により、処理時間の平準化が図れたことも大きなポイントです。以前発生していた定期的な処理の30分以上の所要時間のブレについても移行後は2分以内に収まっています」

圧縮率の大幅向上も松原氏が注目するポイントの一つ。以前と比べ、圧縮後のデータ量は1/5になったという。

将来：

Snowflakeの各種機能で一層のデータ管理工数削減を追求

以前からの課題だったプロジェクト別、事業部門別の分析コスト把握も大きな効果の一つだ。その意義を石井氏はこう説明する。

「以前の一定量の計算リソースが使い放題という環境と比べると、Snowflake移行によりコストは高くなるかもしれませんが、しかし私たちに、この仕組みが見合っていると強く感じています。データ分析コストが見えにくい状況では、問題があるクエリ発行が改善できず、それがレスポンスの遅延にもつながり、結果として『IDAPは使えない』という印象を持たれることを我々はなにより懸念していました。コスト可視化はこれらの課題解決に大きな役割を果たすと考えています」

同社が今後の展開として挙げるのは、データ管理や処理に必要な工数・コストの一層の削減だ。

「データ基盤の巨大化は、管理の困難さにもつながっています。こうした中、私たちが注目するのは、データメッシュによる管理の分散化と異なるデータソースの効果的な統合の実現です。また利用頻度は高くないが利用時には処理速度が求められる大規模なデータについてはIceberg Tableによる保有を検討しています」

Snowflakeについて

Snowflakeは、シンプルかつ効率的で信頼性の高いエンタープライズAIを実現します。Snowflakeのデータクラウドは、世界最大規模の数百の企業を含む世界中の数多くのお客様に利用されており、データ共有、AI/機械学習アプリケーションの構築、ビジネスの強化に貢献しています。これからは、エンタープライズAIの時代です。詳しくは、snowflake.com/ja/ (ニューヨーク証券取引所:SNOW)をご覧ください。