

数兆レコードの人流データをSnowflakeに移行 柔軟なスケーリングとニアゼロメンテナンスにより 社内の誰もがデータにアクセスできる環境を実現

情報通信業



企業 株式会社Agoop
所在地 東京都渋谷区神宮前3-35-8
ハニービル青山 6F

ご利用のSnowflakeワークロード



マルチキャリアのスマートフォンアプリから位置情報を収集し、流動人口データとして提供する株式会社Agoopでは、数千万MAUの位置情報を分析する基盤運用が以前より大きな課題だった。スケールアウトによる対応が難しく、日々増減するデータ処理量への対応ストレスは、気付くと最小限のスタッフ以外はアクセスがしづらい状況につながっていた。スケーリングの柔軟さやニアゼロメンテナンスという特長を備えるSnowflakeへの移行は、この課題を大幅に改善。社内の誰もがアクセスでき、新たなサービス提供への多様な試みを可能にする基盤に生まれ変わった。

このストーリーのハイライト:

- ・スムーズなスケールアウトでパフォーマンスを大幅改善
- ・ニアゼロメンテナンスで運用ストレスを大幅に削減
- ・負荷を意識することなく社内の誰もがアクセスできる基盤の実現

課題:

膨大なデータを処理する分析基盤の スムーズな運用が大きな課題に

観光振興や店舗開発、都市整備のDXにおいて大きな役割を果たしているのが、同意を得たユーザーのスマホアプリから収集される位置情報データを、秘匿加工を行い、人の動きを可視化した流動人口データである。ソフトバンクの基地局再編を支えたデータサイエンティスト部門が分社化して2009年に設立された株式会社Agoopは、人流データの収集・分析を通して、企業や自治体のDXを支援する企業だ。第一の強みと言えるのが、マルチキャリアで収集される数千万MAUの行動データの存在である。また、人流分析の精度を決定するメッシュサイズの細かさや強みの一つで、5m四方という業界最小サイズの人流解析基盤を実現している。取締役 兼 CTO 技術開発本部 部長の加藤 有祐氏はその意義をこう説明する。

「プライバシー観点から、平均滞在人口などをメッシュ単位で提供する際には、50mメッシュで提供していますが、我々が目指す高精度な人流分析において5m四方のメッシュサイズは大きな意味を持ちます。つい先日リ

リースされた、道路一本ごとの人流を分析する新サービス『トラカン』でも新基盤は大きな役割を果たし、従来の人手による目視調査を不要にする同サービスは大きな反響を得ています」

その一方で、以前から求められてきたのが、数兆レコードのデータ量、数千テーブルのデータ分析基盤のスムーズな運用の実現だった。

2015年以来運用してきた同社のデータ分析基盤は、大きく4つの課題に直面していた。一つはパフォーマンスである。

“ 「分析基盤には日々人流データが取り込まれ、日次処理を行います。データ量は日々増減するため、運用管理の観点では処理量に応じてリソースの確保が必要ですが、以前の環境ではスペック変更で1日以上時間が掛かり、その間は作業を中止する必要があるため、事実上スケールアウトによる対応が困難な状況が続いていました」

この問題は結果として、処理遅延の懸念や、サーバーリソースに余裕がある時間帯にアドホック分析の処理を行う必要があるなど、特にエンジニアに対して負担になっていたという。

次にコスト面の課題である。あらかじめ必要なリソースを予約するリザーブドインスタンスという考え方は、リソースをより低コストで調達することを可能にする一方、ピークに合わせて予約する関係上、リソースの無駄が避けられない。同社の試算では、確保したリソースの60%が無駄になっていたという。

3つ目がメンテナンスに関する課題だ。

「当時最も苦労したのは、バージョンアップなどに伴う再起動への対応でした。再起動スケジュールの案内はかなり早くからあり、リスケジュールも可能なのですが、高頻度で求められるバージョンアップへの対応はエンジニアにとってもデータサイエンティストにとっても大きな負担になっていました」

最後の課題が、こうした問題が同社サービスの基盤でもある人流分析そのものを阻害している現状だった。これを加藤氏は「心理的負担」という言葉で表現する。

「一つのトラブルがビジネスへの影響に直結するようなセンシティブな仕

1/7

メッシュ基盤移設に必要な
処理日数を1/7に削減

28%

メッシュ基盤移設コストを約3割削減

40%

ストレージを4割削減

組みであれば、やはり運用の当事者以外は距離を置こうと考えますよね。当社の分析基盤はまさにそうしたエンジニアにとって大きな心理的負担がかかる基盤と言うほかない状況に陥っていました」

解決策：

経営サイドとエンジニアが

同じタイミングでSnowflakeに注目

こうしたなか出会ったのが、Snowflakeのデータ分析基盤だった。

「我々経営層がSnowflakeからアプローチを受けたのと当社のデータエンジニアチームのリーダーがSnowflakeを独自にキャッチしたのはほぼ同時期のことだったようです。彼に私が『おもしろそうなデータ基盤があるんだけど、どう思いますか?』と投げかけたところ、『実は僕らも今、同じことを提案しようと思っていたんです』という返答があり、そこから一気に移行に向けた動きが進むことになりました」

Snowflake移行に向けた検証がスタートしたのは2022年12月のこと。検証は加藤氏とデータベースエンジニア2名、データサイエンティスト1名という最小限の体制で行われている。その背景には、検証プロセスにおいてさまざまなことを迅速に試す上では最小限の体制が望ましいという加藤氏の考え方がある。

また検証パートでは、Snowflakeへの移行効果が学習コストや移行コストを上回るか否かを入念に確認したという。

「まず注目したのは、これまでデータ分析を行ってきたクエリがSnowflakeに対応するのかという点でした。調査すると、すでにクエリ変換ツールが広く普及し、移行済みのユーザー企業によるレビューも豊富である上、Snowflakeが提供するドキュメントも非常に理解しやすいものでした。そういう意味で、学習コストに関する障壁は極めて低いと判断しました」

同社は2022年4月からSnowflakeへの移行を開始し、同年8月には実運用を開始している。

結果：

スムーズなスケールアウトが

多くの課題を解決

移行効果としてまず注目したいのが、かねての課題だったパフォーマンス面の効果だ。

「まず注目したいのはスケールアウトがわずか数秒で、作業を中断することなく行えるようになった点です。それにより、日々の処理量の増減でエンジニアが頭を悩ますことも、アドホック分析の時間を調整する必要もなくなりました。またパフォーマンスの観点では、クエリのボトルネックを特定する、クエリプロファイルも大きな役割を果たしています。大きなウェアハウスを動かす際や、クエリのレビューなどで積極的に活用しています」

スケールアウトが自在に行えるようになったことは、コスト改善にもつながっている。その分かりやすい例がメッシュサイズの細密化の実現である。

「以前当社は、50mというメッシュサイズで人流データ基盤を作成していました。これ自体、業界最小サイズなのですが、道路レベルの人流把握の実現にはより精緻なデータが求められます。以前から5mメッシュサイズへの移行の検討を進めていたのですが、以前のデータ分析環境で試算したところ、それには約200日の処理時間が必要になり、一度は計画を見直したというのが正直なところでした。しかしSnowflakeへの移行により、29日の処理時間でそれが実現できています。また、コストは28%削減、ストレージ容量も40%の削減と、処理日数、コスト、ストレージ容量全てにおいて大きな効果がありました」

Snowflakeが掲げるニアゼロメンテナンスの効果にも注目したい。

「定期メンテナンスについては完全にバックグラウンドで行われるので、処理を中断し再起動する手順を考える必要はなくなりました。同様にストレージ上限を気にする必要がなくなったことで、我々はフルマネージドに近い感覚でデータ基盤を運用できるようになっています」

さらに大きいのが、同社のビジネスの基盤である同意を得たユーザーのスマートフォンアプリから収集される位置情報データを一元管理し、負荷を気にすることなくアクセスできるデータ基盤が実現したことが持つ意義だ。

「センシティブなデータ基盤からできるだけ距離を置きたいという意識が全面的に書き換えられたことは、個人的にはSnowflake移行の最大の効果だったと考えています。すでに全従業員のうち7割程度がSnowflakeにアクセスし、これまでは考えられなかったような部門横断型のプロジェクトもすでにスタートしています」

将来：

マーケットプレイスのデータ共有による 人流データの一層の活用に期待

人流データは単体での利用のほか、天候データやPOSデータと掛け合わせることで新たな価値を創出する。すでに同社の顧客の中には、こうした取り組みを開始する企業も現れはじめているという。

こうした中、同社が2023年2月に開始したのが、Snowflakeマーケットプレイスを利用した人流データのサンプルデータ提供である。

「人流データ提供はビジネスの柱の一つとしてこれまでも行ってきましたが、新たなデータ提供には、顧客内での利用に際しての社内調整や各種契約手続きなど、どうしても提供までに時間がかかっていました。それを考えると、わずか数秒でデータ共有を実現するマーケットプレイスの意義は極めて大きいと思います。コストやデータ量の問題もあり、これまで人流データ活用には大きな制約がありましたが、マーケットプレイス活用はその普及に大きな役割を果たすと考えています」

Snowflakeについて

Snowflakeは、あらゆる組織がSnowflakeデータクラウドを用いて自らのデータを最大限に活用するのを支援します。多くのユーザー企業がデータクラウドを利用して、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、データアプリケーションの推進、さらには多様なAI/MLや分析ワークロードの実行を進めています。データやユーザーがどこに存在するかに関係なく、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。多くの業界の何千社もの企業（2023年7月31日時点で、2023年Forbes Global 2000社（G2K）のうち639社を含む）が、Snowflakeデータクラウドを全社で幅広いビジネスに活用しています。

詳しくは、[snowflake.com](https://www.snowflake.com)をご覧ください。