

プロジェクト実装にかかるリードタイムとコストを効率化。 継続的な事業拡大に向けて大きなベネフィット。



株式会社ジオロジックは、位置情報データに特化したインターネット広告を提供する、最先端のアドテクノロジー企業である。高い技術力をバックボーンに、広告関連商品を開発して販売している。同社の設立は2014年11月であり、そこからわずか数年で広告主は1500社以上、広告代理店は数百社以上と取り引きをするまでに成長している。アドテクノロジー業界を強かにリードしている企業と言える。リアルタイムの位置情報データを駆使する同社がメインターゲットにしているのは、地域指定で広告を出すエリアマーケティングの分厚いマーケットである。従来はチラシなどでカバーされていた領域でもあり、デジタル広告としては新市場ともいえる分野で、大きな成長が見込まれている。クライアントには、不動産業や学習塾、カーディーラーなど、地域に拠点を持つ大手企業が多い。大手マンションディベロッパーのほとんどと取り引きがあるなど、同社は高関与商品を取り扱う大企業の地域密着型広告でめざましい成果をあげている。



株式会社ジオロジック 取締役CTO 藪本 晃輔氏

採用したプロダクト



ハイライト

コストを**20%削減!**
将来的には
30~40%の削減を
予想!

3~4日かかっていた
ジョブフローが
2時間以内に!

攻めの
プロダクト開発が
可能に!

課題

データプロジェクトを進める上で、
時間・柔軟性・パフォーマンスに課題。

これまで、Amazon Redshiftを中心としたプラットフォーム上で、データプロジェクトを進めていたが、事業拡大に伴い、幾つかの課題に直面していた。1つ目の課題は、必要なデータを抽出するための時間がかかりすぎて、ビジネス側の要件にタイムリーに対応できない点であった。例えば、ある

特定の位置から半径1km以内のターゲットに広告を配信したい、という要件があった場合、対象となるデータセットの抽出にかなりの時間を要していた。2つ目の課題は、当時の環境の技術的な制限により、積極的な営業活動や製品開発に転じることができなかった点である。例えば、AWS Redshiftは、

データのソートする順番やIDの定義の変更をした場合、テーブルを作り直す必要がある等、DDL (Data Definition Language)の柔軟性に欠けていると感じていた。UPSERTは機能としてサポートしているが、同社のデータサイズでは、パ

フォーマンスに課題があり、一方でINSERTではコストが肥大化してしまうという状況であった。また、SQL関数も限られており、特に機械学習向けの集計関数が多く用意されていない点も懸念であった。

比較検討

現状のAWS上に構築可能であること、基盤構築に工数がかからないこと、そして、プロジェクトの実装時間を短縮できることが採用の決め手。

当初、AWS GlueやAmazon EMRなどを活用し、AWSプラットフォーム上での課題解決を試みた。しかし、AWS Glueはストリーミング処理の実装ができず、AWS EMRはUPSERT処理ができないことが判明した。いずれも今回のユースケースでは必須である。また、担当メンバーは、HadoopやSparkなどの解析基盤をゼロから作った経験が無く、基盤構築のために工数をかけるよりは、より手軽でコストのかからない方法が理想的であった。そして、GlueやEMRよりも、Databricksの方が使

いやしく、ヘルプページやサポートも充実していた。また、GCP (Google Cloud Platform)も比較検討を行ったが、AWSからデータ転送をしなければいけない点が大きなネックなると判断した。結果として、現状のAWS上に構築可能であること、基盤構築に工数がかからないこと、そして、プロジェクトの実装時間を短縮できること、これら3点の理由から、Databricksの採用を決定した。

ビジネスインパクト

プロジェクト実装にかかるリードタイムが大幅に短縮。さらに、約20%のコスト削減を実現。

Databricksを採用することで、いくつかの明確な導入効果が確認できている。1つ目の導入効果は、プロジェクト実装にかかるリードタイムが大幅に短縮することが可能となった点である。これまで、大量の過去の位置情報から、ある特定の情報を引き出すような場合、Redshiftで3~4日かかっていたが、Databricksでは数時間に入手できるようになった。2つ目の導入効果は、約20%のコスト削減である。現状では一部のRedshift環

境の縮退に留まっているが、さらにプロジェクトを進めて行った場合、総合的に30~40%のコスト削減を確信している。3つ目の導入効果は、データ保存はAmazon S3、解析基盤はDatabricksとすることによって、データプロジェクトのコストがデータ量の増加に比例しない環境になった点である。これは、継続的な事業拡大を前提としている会社にとって、非常に大きなベネフィットであると言える。

今後の期待

外部への監視ツールとの利便性、連携、強化に期待。

外部への監視ツールへの統合がもう少し楽にできると良いと感じていた。これに関して、近日中にリリース予定であるSpark 3.0によって、これまでサポートしていたGangliaに加え、PrometheusやGraphiteなどの追加され、より利便性が

向上する予定だ。また、IDEとNotebookとの連携にはいくつかの煩わしいオペレーションが求められる。これに関して、近日リリース予定のWorkspace 2.0において、Git プロジェクト毎の連携が簡単になる予定である。