

データ ウェアハウス構築に BigQuery を採用することで、5～10 倍のパフォーマンスと運用性、利便性を向上



freee株式会社

<https://www.freee.co.jp/>
〒141-0032 東京都品川区大崎 1-2-2
アートヴィレッジ大崎セントラルタワー 21F

2012年7月設立。「スモールビジネスを、世界の主役に。」というミッションに基づき、個人/法人が利用できるクラウド型の会計ソフト「freee会計」を中核に、人事労務、申告、マイナンバー管理、会社設立、開業で構成される統合型経営プラットフォームの開発、提供を事業として展開。現在、有料課金ユーザー企業数は、約38万事業所(2022年6月末現在)。2021年より、「統合型コーポレートカード」「freeeカード Unlimited」を提供することで、ファイナンス面からもスモールビジネスを支援しています。

■インタビュー(写真左から)

- ・プロダクト基盤本部 技術基盤部 マネージャ 今岡 久敏 氏
- ・プロダクト基盤本部 技術基盤部 中道 鋪彌 氏

個人事業や中小企業などのスモール ビジネスに携わるすべての人が、自由に自然体で経営できる環境をつくるための「統合型経営プラットフォーム」を開発、提供する freee株式会社(以下、freee)。2013年3月にクラウド会計ソフト「freee」をリリースすると同時に、エンジニア主導でデータ基盤を活用し、ユーザーの事業拡大を支援してきました。このデータ基盤が10年を経て、パフォーマンスや運用性、利便性に課題が生じたことから刷新を決定。新しいデータ基盤の中核に BigQuery が採用されています。このプロジェクトについて、プロダクト基盤本部の2名に話を伺いました。

■利用しているサービス

BigQuery、Cloud Storage、VPC Service Controls、Access Context Manager、Looker Studio、コネクテッド シート、Colaboratory、Dataproc、Firebase、Google アナリティクス

■利用しているソリューション

スマート アナリティクス

データ分析チームが小規模に使っていた実績が BigQuery 採用の決め手の1つに

統合型クラウド ERP、オープン プラットフォーム、ユーザー ネットワークの3つの柱で構成される「統合型経営プラットフォーム」により、スモール ビジネスをバック オフィス業務の手間から解放し、誰もが自由に、自然体で経営できるようにすることを目指す freee。業務支援の一環として、データ ウェアハウスを構築し、さまざまなデータ分析を行い、その結果をビジネスの拡大に生かしてきました。プロダクト基盤本部 技術基盤部 中道 鋪彌氏は、「データ ウェアハウスは、ARR(年間経常収益)や解約率など、SaaS ビジネスで把握しておきたい事業 KPI の可視化に利用していました。またサポート部門が、データ ドリブでスピーディーにサポート活動の改善箇所を特定したり、エンジニアがデプロイの回数やチケットの内容の分析をすることで生産性の改善にも役立てています。」と話します。しかし、このデータ ウェアハウスは、構築から約10年が経過したことから、パフォーマンスの後退、高い運用負荷、利便性の欠如という3つの課題を抱えていました。例えば、複雑なクエリを実行すると、結果が1日返ってこないこともあり、分析者の集中力を大幅に削ぐ要因になっていました。クエリ実行の待ち時間をいかに短縮するかを大きな課題として、プラットフォームから見直すことにしました。また、従来のデータ ウェアハウスは、パフォーマンスを引き出すために、インデックスの設計やチューニングなどの職人芸が必要だったため、運用負荷が高く

なっており、改善が求められていました。さらに、全社的に利用していたインターフェースは、オープン ソースの BI ツールである Redash(リダッシュ)だけだったため、データ アクセスの手段を増やすことで、データ基盤の利便性をより一層向上させることも必要でした。

プロダクト基盤本部 技術基盤部 今岡 久敏氏は、「ユーザーが、データの格納構造を意識せずに複雑なクエリを実行した場合、想定以上のデータを抽出し、集計に時間がかかっていました。高速化の努力は続けてきましたが、改善も限界にきていました。大きなクエリが1つ実行されると、ほかのクエリも時間がかかってしまうという状況でした」と話します。

そこで2020年10月より、本格的なデータ ウェアハウス刷新の検討を開始。社内の調整や承認を経て、2021年4月に BigQuery の採用を決定します。最大の決め手は、3つの課題(パフォーマンスの後退、高い運用負荷、利便性の欠如)の解決が期待できることでした。中道氏は、「ほかにもいくつかのデータ ウェアハウスを検討しましたが、以前からデータ分析チームが小規模ではあるものの BigQuery を使っていた実績があったことも採用の決め手でした。現場からも BigQuery の利便性とパフォーマンスを推す声があり、BigQuery を採用するほうが現場のモチベーションも上がると考えました」と話しています。

BigQuery の外部データソースの仕組みで移行のコストを大幅に削減

BigQuery を採用した新しいデータ ウェアハウスは、free の各サービスが稼働している他社のクラウド環境から、ログやセールス情報、アプリケーション データなど、分析に必要なデータを同クラウド上に構築した Spark ETL で抽出・加工した後、BigQuery に転送し、Redash などの周辺サービスでデータ分析を行う仕組みになっています。ログは 1 時間に 1 回、そのほかのデータは 1 日に 1 回のサイクルでデータが転送されています。

今回の開発について今岡氏は、「BigQuery に関しては、以前はデータ分析の専門家のツールというイメージでした。しかし今回、システムを構築している最中に、データ利用者の視点でも管理者の視点でも、あったらいいのと思っていた機能がどんどん増えてきたので、現在は誰もが簡単に使えるデータ基盤としてフレンドリーになったと実感しています」と話します。

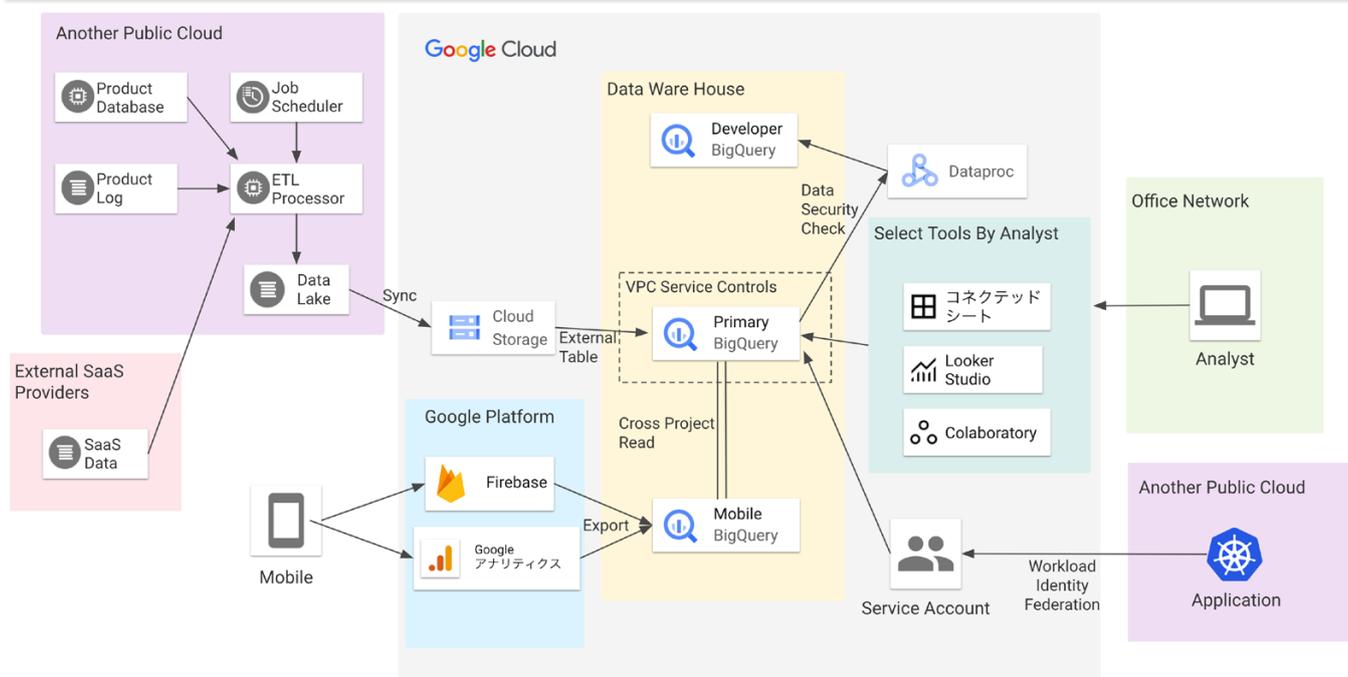
データ基盤刷新プロジェクトは、エンジニアが延べ 10 名弱、開発期間は 1 年半程度でした。2020 年 10 月より本格的な検証を開始し、2021 年 4 月に BigQuery へのデータ転送を開始。7 月から旧システムとの並行稼働を開始し、一部の利用者向けに公開して、10 月に全社公開しています。2022 年 1 月には、旧システムを廃止していますが、現場の反発もなく、スムーズな移行を実現しています。

BigQuery を採用した効果を今岡氏は、「監査ログの充実、BigQuery のメリットの 1 つで、どのテーブル、どのカラムを利用しているかを SQL 文の分析なしで把握できるのは大きなメリットでした。また移行に関しては、オープン フォーマットのデータを Cloud Storage 上に置く外部データソースの仕組みが、既存のデータレイクの方式にマッチしていて、既存データをそのまま BigQuery に移行できました。これにより、移行のコストを大幅に削減できました」と話します。

BigQuery を導入したことで、当初の目的だったパフォーマンスの後退、高い運用負荷、利便性の欠如の 3 つの課題解決を実現しています。

中道氏は、「パフォーマンス面では、旧環境とのクエリ実行時間を比較、検証した結果、BigQuery の方が 5 ~ 10 倍ほどクエリを高速に処理できました。運用面では、BigQuery はインデックスの設計やチューニングなどの職人芸が不要なので、運用の効率化が可能になり、より価値を創出する業務にリソースを集中することができています。また利便性の向上では、Looker Studio やコネクテッド シート、Collaboratory など、BigQuery の充実した周辺サービスやエコシステムを利用できるようになりました」と話しています。

Architecture



今後、BigQuery と Apache Spark を連携する機能にも期待

BigQuery を採用したデータ ウェアハウスの構築における Google Cloud のサポートについて今岡氏は、次のように話します。「移行の初期段階で、技術的なサポートをしてもらえたので非常に助かりました。今後の要望としては、BigQuery は進化が激しいので、アップデート情報をタイムリーに提供してもらえると助かります。それ以外に関しては、本当に満足しています。」

また中道氏は、「早い段階で、VPC Service Controls と Access Context Manager を組み合わせた運用を決めていたのですが、当時はこの組み合わせではコネクテッドシートが利用できないという課題がありました。そこで Google Cloud の担当者

に要望を伝えたところ、思った以上に短期間で対応してもらえたのでありがたかったです」と話します。

今後の展望について今岡氏は、「現在、データ分析基盤として Apache Spark をかなり使っているので、今後 BigQuery と Apache Spark を連携する機能を利用してみたいと思っています。また機械学習に関しては、データ分析基盤として、いかに容易に利用できるようにするか、ストレスなくデータ分析ができるようになるかを検討しています。」と話しています。

Google Cloud を活用することで、ビジネスの将来に注力できるようになります。インフラストラクチャの管理やサーバーのプロビジョニング、ネットワークの構成などに起因する負担を軽減することができます。つまり、インベーターもプログラマーも、自分の本来の仕事に集中することができます。

お問い合わせはこちらから
<https://goo.gl/CCZL78>

